

References

- 1) R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith and J. C. Venter, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science*, **269**(5223), 496-512(1995).
- 2) International Human Genome Sequencing Consortium (E. S. Lander, *et al.*), Initial sequencing and analysis of the human genome, *Nature*, **409**(6822), 860-921(2001).
- 3) J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R.-R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang,

X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guig?, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yoosheph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, Pa. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu, The Sequence of the Human Genome, *Science*, **291**(5507), 1304-1351(2001).

- 4) J. G. Lawrence and H. Ochman, Molecular archaeology of the *Escherichia coli* genome, *Proc. Natl. Acad. Sci. USA*, **95**(16), 9413-9417(1998).
- 5) L. B. Koski, R. A. Morton and G. B. Golding, Codon bias and base composition are poor indicators of horizontally transferred genes, *Mol. Biol. Evol.*, **18**(3), 404-412(2001).
- 6) P. M. Sharp and W. H. Li, The codon Adaptation Index—a measure

- of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.*, **15**(3), 1281-1295(1987).
- 7) R. F. Doolittle, Biodiversity: microbial genomes multiply, *Nature*, **416**(6882), 697-700(2002).
 - 8) J. Hacker, G. Blum Oehler, I. Muhldorfer and H. Tschape, Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution, *Mol. Microbiol.*, **23**(6), 1089-1097(1997).
 - 9) R. T. Okinaka, K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson and P. J. Jackson, Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes, *J. Bacteriol.*, **181**(20), 6509-6515(1999).
 - 10) T. Kaneko, Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada and S. Tabata, Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, *DNA Res.*, **7**(6), 331-338(2000).
 - 11) E. Morett and P. Bork, Evolution of new protein function: recombinational enhancer Fis originated by horizontal gene transfer from the transcriptional regulator NtrC, *FEBS Lett.*, **433**(1-2), 108-112(1998).
 - 12) R. L. Tatusov, D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova and E. V. Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.*, **29**(1), 22-28(2001).
 - 13) I. Yanai, Y. I. Wolf and E. V. Koonin, Evolution of gene fusions: horizontal transfer versus independent events, *Genome. Biol.*, **3**(5), research0024.1- research0024.13(2002).
 - 14) S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**(17), 3389-3402(1997).
 - 15) <ftp://ncbi.nlm.nih.gov/genbank/genomes/bacteria/>

- 16) Y. Kawarabayasi, Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jinno, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, K. Kubota, Y. Nakamura, N. Nomura, Y. Sako, and H. Kikuchi, Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1, *DNA Res.*, **6**(2), 83-101(1999).
- 17) H. P. Klenk, R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. D'Andrea, C. Bowman, C. Fujii, S.A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese and J. C. Venter, The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, *Nature*, **390**(6658), 364-370(1997).
- 18) C. J. Bult, O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen and J. C. Venter, Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, *Science*, **273**(5278), 1058-1073(1996).
- 19) D. R. Smith, L. A. Doucette-Stamm, C. Deloughery, H.-M. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, H. Safer, D. Patwell, S. Prabhakar, S. Mcdougall, G. Shimer, A. Goyal, S. Pietrokovski, G. M. Church, C. J. Daniels, J.-I. Mao, P. Rice, J. Nolling, and J. N. Reeve, Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics, *J. Bacteriol.*, **179**(22), 7135-7155(1997).

- 20) Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki and H. Kikuchi, Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3, *DNA Res.*, **5**(2), 55-76(1998).
- 21) <http://www.genoscope.cns.fr/Pab/>
- 22) G. Deckert, P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen and R. V. Swanson, The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*, *Nature*, **392**(6674), 353-358(1998).
- 23) F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S.-Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M.-F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S.-M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O' Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S.-H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B.-S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane,

- K. Yasumoto, K. Yata, K. Yoshida, H.-F. Yoshikawa, E. Zumstein, H. Yoshikawa and A. Danchin, The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*, *Nature*, **390**(6657), 249-256(1997).
- 24) C. M. Fraser, S. Casjens, W. M. Huang, G. G. Sutton, R. A. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. K. Hickey, M. Gwinn, B. Dougherty, J. F. Tomb, R. D. Fleischmann, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van-Vugt, N. Palmer, M. D. Adams, J. D. Gocayne, J. Weidman, T. Utterback, L. Waththey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fujii, M. D. Cotton, K. Horst, K. Roberts, B. Hatch, H. O. Smith and J. C. Venter, Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*, *Nature*, **390**(6660), 580-586(1997).
- 25) S. Kalman, W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis and R. S. Stephens, Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*, *Nat. Genet.*, **21**(4), 385-389(1999).
- 26) R. S. Stephens, S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin and R. W. Davis, Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*, *Science*, **282**(5389), 754-759(1998).
- 27) O. White, J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, K. S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J. J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K. S. Makarova, L. Aravind, M. J. Daly, K. W. Minton, R. D. Fleischmann, K. A. Ketchum, K. E. Nelson, S. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, Genome sequence of the radioreistant bacterium *Deinococcus radiodurans* R1, *Science*, **286**(5444), 1571-1577(1999).
- 28) F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao, The complete genome sequence of *Escherichia coli* K-12., *Science*, **277**(5331), 1453-1474(1997).

- 29) J.-F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Wathley, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser and J. C. Venter, The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature*, **388**(6642), 539-547(1997).
- 30) R. A. Alm, L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis and T. J. Trust, Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*, *Nature*, **397**(6715), 176-180(1999).
- 31) C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison III, and J. C. Venter, The minimal gene complement of *Mycoplasma genitalium*, *Science*, **270**(5235), 397-403(1995).
- 32) R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkel, B. C. Li and R. Herrmann, Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*, *Nucleic Acids Res.*, **24**(22), 4420-4449(1996).
- 33) S. G. Andersson, A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler and C. G. Kurland, The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria, *Nature*, **396**(6707), 133-140(1998).
- 34) T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto, T. Kimura, T.

- Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata, Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, *DNA Res.*, **3**(3), 109-136(1996).
- 35) K. E. Nelson, R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*, *Nature*, **399**(6734), 323-329(1999).
- 36) C. M. Fraser, S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, C. Fujii, S. Garland, B. Hatch, K. Horst, K. Roberts, M. Sandusky, J. Weidman, H. O. Smith and J. C. Venter, Complete genome sequence of *Treponema pallidum*, the syphilis spirochete, *Science*, **281**(5375), 375-388(1998).
- 37) S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, **393**(6685), 537-544(1998).
- 38) <ftp://ftp.ncbi.nih.gov/blast/db/>
- 39) W. R. Pearson and D. J. Lipman, Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. USA*, **85**(8), 2444-2448(1988).

- 40) J. D. Thompson, D. G. Higgins and T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**(22), 4673-4680(1994).
- 41) M. Riley, Systems for categorizing functions of gene products, *Curr. Opin. Struct. Biol.*, **8**(3), 388-392(1998).
- 42) M. Ingelman, V. Bianchi and H. Eklund, The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7Å resolution, *J. Mol. Biol.*, **268**(1), 147-157(1997).
- 43) C. J. Harrison, M. Hayer-Hartl, M. Di, Liberto, F. Hartl and J. Kuriyan, Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK, *Science*, **276**(5311), 431-435(1997).
- 44) M. Ingelman, S. Ramaswamy, V. Niviere, M. Fontecave and H. Eklund, Crystal structure of NAD(P)H:flavin oxidoreductase from *Escherichia coli*, *Biochemistry*, **38**(22), 7040-9(1999).
- 45) H. S. Subramanya, L. K. Arciszewska, R. A. Baker, L. E. Bird, D. J. Sherratt and D. B. Wigley, Crystal structure of the site-specific recombinase, XerD, *EMBO J.*, **16**(17), 5178-5187(1997).
- 46) M. M. Thayer, H. Ahern, D. Xing, R. P. Cunningham and J. A. Tainer, Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure, *EMBO J.*, **14**(16), 4108-4120(1995).
- 47) T. Kawashima, C. Berthet-Colominas, M. Wulff, S. Cusack and R. Leberman, The structure of the *Escherichia coli* EF-Tu·EF-Ts complex at 2.5Å resolution, *Nature*, **379**(6565), 511-518(1996).
- 48) <http://genolist.pasteur.fr/Colibri/>
- 49) <http://www.rcsb.org/pdb/>
- 50) <http://opal.biology.gatech.edu/GeneMark/>
- 51) C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells and J. M. Thornton, CATH—a hierarchic classification of protein domain structures, *Structure*, **5**(8), 1093-1108(1997).
- 52) <http://spiral.genes.nig.ac.jp/homology/>

- 53) C. Médigue, T. Rouxel, P. Vigier, A. Hénaut and A. Danchin, Evidence for horizontal gene transfer in *Escherichia coli* speciation, *J. Mol. Biol.*, **222**(4), 851-856(1991).
- 54) M. Borodovsky, J. D. McIninch, E. V. Koonin, K. E. Rudd, C. Médigue and A. Danchin, Detection of new genes in a bacterial genome using Markov models for three gene classes, *Nucleic Acids Res.*, **23**(17), 3554-3562(1995).
- 55) M. I. Bellgard and T. Gojobori, Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content, *Gene*, **238**(1), 33-37(1999).
- 56) A. Muto and S. Osawa, The guanine and cytosine content of genomic DNA and bacterial evolution, *Proc. Natl. Acad. Sci. USA*, **84**(1), 166-169(1987).
- 57) G. Bernardi and G. Bernardi, Codon usage and genome composition, *J. Mol. Evol.*, **22**(4), 363-365(1985).
- 58) F. Bourgoïn, A. Pluvinet, B. Gintz, B. Decaris and G. Guedon, Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci?, *Gene*, **233**(1-2), 151-161(1999).
- 59) Y. Ohfuku, H. Tanaka and M. Uebayasi, Relationship between nucleotide sequence and 3D protein structure of six genes in *Escherichia coli*, by analysis of DNA sequence using a Markov model, *Chem-Bio Informatics J.*,(2002), in press.
- 60) A. C. R. da Silva, J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. C. Alves, A. M. do Amaral, M. C. Bertolini, L. E. A. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chambergó, L. P. Ciapina, R. M. B. Cicarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El-Dorry, J. B. Faria, A. J. S. Ferreira, R. C. C. Ferreira, M. I. T. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. M. Lemos, M. V. F. Lemos, E. C. Locali, M. A. Machado, A. M. B. N. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. M. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. M. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. D. Sena, C. Silva, R. F. de Souza, L. A. F. Spinola, M. A. Takita, R. E. Tamura,

- E. C. Teixeira, R. I. D. Tezza, M. Trindade dos Santos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal and J. P. Kitajima, Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities, *Nature*, **417**(6887), 459-463(2002).
- 61) J. M. Isas, S. M. Yannone and B. K. Burgess, *Azotobacter vinelandii* NADPH:ferredoxin reductase cloning, sequencing, and overexpression, *J. Biol. Chem.*, **270**(36), 21258-21263(1995).
- 62) C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook, Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory and M. V. Olson, Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature*, **406**(6799), 959-964(2000).
- 63) A. J. G. Simpson, F.C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. C. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bové, M. R. S. Briones, M. R. P. Bueno, A. A. Camargo, L. E. A. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. R. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J. S. Ferreira, V. C. A. Ferreira, J. A. Ferro, J. S. Fraga, S. C. França, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. S. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. C. Leite, E. G. M. Lemos, M. V. F. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. B. N. Madeira, H. M. F. Madeira, C. L. Marino, M. V. Marques, E. A. L. Martins, E. M. F. Martins, A. Y. Matsukuma, C. F. M. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteiro-Vitorello, D. H. Moon, M. A. Nagai, A. L. T. O. Nascimento, L. E. S. Netto, A. Nhani Jr, F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. De Oliveira, R. C. De Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. G. Pereira, H. A. Pereira Jr, J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J. De M. Rosa, V. E. De Rosa Jr, R. G. De S?, R. V. Santelli, H. E. Sawasaki, A. C. R. Da Silva, A. M. Da Silva, F. R. Da Silva,, W. A. Silva, J. F. Da Silveira, M. L. Z. Silvestri, W. J. Siqueira, A. A. De Souza, A. P. De Souza, M. F. Terenzi, D. Truffi, S. M. Tsai,

- M. H. Tsuhako, H. Vallada, M. A. Van Sluys, S. Verjovski-Almeida, A. L. Vettore, M. A. Zago, M. Zatz, J. Meidanis and J. C. Setubal, The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature*, **406**(6792), 151-157(2000).
- 64) T. Shigematsu, S. Hanada, M. Eguchi, Y. Kamagata, T. Kanagawa and R. Kurane, Soluble methane monooxygenase gene clusters from trichloroethylene-degrading *Methylobacter* sp. strains and detection of methanotrophs during in situ bioremediation, *Appl. Environ. Microbiol.*, **65**(12), 5198-5206(1999).
- 65) M. Salanoubat, S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel, Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Signier, P. Thebault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach and C. A. Boucher, Genome sequence of the plant pathogen *Ralstonia solanacearum*, *Nature*, **415**(6871), 497-502(2002).
- 66) G. Sridhar, Prasad, N. Kresge, A. B. Muhlberg, A. Shaw, Y. S. Jung, B. K. Burgess and C. D. Stout, The crystal structure of NADPH:ferredoxin reductase from *Azotobacter vinelandii*, *Protein Sci.*, **7**(12), 2541-2549(1998).
- 67) Y. Komeiji, M. Uebayasi, R. Takata, A. Shimizu, K. Itsukashi and M. Taiji, Fast and accurate molecular dynamics simulation of a protein using a special purpose compute, *J. Comp. Chem.*, **18**, 1546-1563(1997).
- 68) M. N. Burtnick and D. E. Woods, Isolation of polymyxin B-susceptible mutants of *Burkholderia pseudomallei* and molecular characterization of genetic loci involved in polymyxin B resistance, *Antimicrob. Agents Chemother.*, **43**(11), 2648-2656(1999).
- 69) S. C. Winans and G. C. Walker, Conjugal transfer system of the IncN plasmid pKM101, *J. Bacteriol.*, **161**(1), 402-410(1985).
- 70) A. M. Brown and N. S. Willetts, A physical and genetic map of the IncN plasmid R46, *Plasmid*, **5**(2), 188-201(1981).
- 71) M. Redenbach, K. Ikeda, M. Yamasaki and H. Kinashi, Cloning and physical mapping of the EcoRI fragments of the giant linear plasmid SCP1, *J. Bacteriol.*, **180**(10), 2796-2799(1998).
- 72) Y. M. Kim and S. N. Lee, direct Submission to NCBI

- 73) X. Peng, T. Egashira, K. Hanashiro, E. Masai, S. Nishikawa, Y. Katayama, K. Kimbara and M. Fukuda, Cloning of a *Sphingomonas paucimobilis* SYK-6 gene encoding a novel oxygenase that cleaves lignin-related biphenyl and characterization of the enzyme, *Appl. Environ. Microbiol.*, **64**(7), 2520-2527(1998).
- 74) P. J. Langer and G. C. Walker, Restriction endonuclease cleavage map of pKM101: relationship to parental plasmid R46, *Mol. Gen. Genet.*, **182**(2), 268-272(1981).
- 75) D. H. Nies, A. Nies, L. Chu and S. Silver, Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*, *Proc. Natl. Acad. Sci. USA*, **86**(19), 7351-7355(1989).
- 76) T. Kunito, T. Kusano, H. Oyaizu, K. Senoo, S. Kanazawa and S. Matsumoto, Cloning and sequence analysis of *czc* genes in *Alcaligenes* sp. strain CT14, *Biosci. Biotechnol. Biochem.*, **60**(4), 699-704(1996).
- 77) C. Rensing, T. Pribyl and D. H. Nies, New functions for the three subunits of the CzcCBA cation-proton antiporter, *J. Bacteriol.*, **179**(22), 6871-6879(1997).
- 78) N. Medina, F. Vannier, B. Roche, S. Autret, A. Levine and S. J. Seror, Sequencing of regions downstream of *addA* (98 degrees) and *citG* (289 degrees) in *Bacillus subtilis*, *Microbiology*, **143** (Pt. 10), 3305-3308(1997).
- 79) M. Gottfert, S. Rothlisberger, C. Kundig, C. Beck, R. Marty and H. Hennecke, Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome, *J. Bacteriol.*, **183**(4), 1405-1412(2001).
- 80) A. C. Martin, R. Lopez and P. Garcia, Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1, *J. Virol.*, **70**(6), 3678-3687(1996).
- 81) M. Redenbach, H. M. Kieser, D. Denapaite, A. Eichner, J. Cullum, H. Kinashi and D. A. Hopwood, A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome, *Mol. Microbiol.*, **21**(1), 77-96(1996).

- 82) T. Iizumi and K. Nakamura, Cloning, nucleotide sequence, and regulatory analysis of the *Nitrosomonas europaea dnaK* gene, *Appl. Environ. Microbiol.*, **63**(5), 1777-1784(1997).
- 83) Y. Izumiya, H. K. Jang, M. Sugawara, Y. Ikeda, R. Miura, Y. Nishimura, K. Nakamura, T. Miyazawa, C. Kai and T. Mikami, Identification and transcriptional analysis of the homologues of the herpes simplex virus type 1 UL30 to UL40 genes in the genome of nononcogenic Marek's disease virus serotype 2, *J. Gen. Virol.*, **80** (Pt. 9), 2417-2422(1999).
- 84) T. A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp. Ser.*, **41**, 95-98(1999).
- 85) J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz and A. M. Weiner, Molecular Biology of the Gene (Fourth Edition), The Benjamin/Cummings Publishing Company Inc., Menlo Park, California, 1987, pp.1148-1149.
- 86) X. Carbonell and A. Villaverde, Insertional mutagenesis in the tail-spike protein of bacteriophage P22, *Biochem. Biophys. Res. Commun.*, **244**(2), 428-433(1998).
- 87) Y. Ohfuku, H. Tanaka and M. Uebayasi, Identification of lateral gene in the genomic DNA sequence, by using the G+C content at the third position of synonymous codons of orthologous genes as an index, *Chem-Bio Informatics J.*, **2**(2), 58-73(2002).
- 88) J. Osipiuk and A. Joachimiak, Cloning, sequencing, and expression of *dnaK*-operon proteins from the thermophilic bacterium *Thermus thermophilus*, *Biochim. Biophys. Acta.*, **1353**(3), 253-265(1997).
- 89) E. C. de Macario, C. B. Dugan and A. J. Macario, Identification of a *grpE* heat-shock gene homolog in the archaeon *Methanosarcina mazei*, *J. Mol. Biol.*, **240**(1), 95-101(1994).
- 90) A. J. Macario and E. C. de Macario, The archaeal molecular chaperone machine: peculiarities and paradoxes, *Genetics*, **152**(4), 1277-1283(1999).
- 91) E. Ikeda, S. Yoshida, H. Mitsuzawa, I. Uno and A. Toh-e, YGE1 is a yeast homologue of *Escherichia coli grpE* and is required for maintenance of mitochondrial functions, *FEBS Lett.*, **339**(3), 265-268(1994).