

Appendix A: References

- ADAMS, M.D., CELNIKER, S.E., HOLT, R.A., EVANS, C.A., GOYAYNE, J.D., AMANATIDES, P.G., SCHERER, S.E., LI, P.W., HOSKINS, R.A., GALLE, R.F., *et al.* 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.
- AKMAN, L., YAMASHITA, A., WATANABE, H., OSHIMA, K., SHIBA, T., HATTORI, M. & AKSOY, S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics* **32**: 402-407.
- ALBRECHT, M., SHARMA, C.M., REINHARDT, R., VOGEL, J. & RUDEL, T. 2010. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Research* **38**: 868-877.
- AL-HASANI, K., BOYCE, J., MCCARL, V.P., BOTTOMLEY, S., WILKIE, I. & ADLER, B. 2007. Identification of novel immunogens in *Pasteurella multocida*. *Microbial Cell Factories* **6**: 3.
- ALLRED, D.R., MCGUIRE, T.C., PALMER, G.H., LEIB, S.R., HARKINS, T.M., MCELWAIN, T.F. & BARBET, A.F. 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in *Anaplasma marginale*. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 3220-3224.
- ALLSOPP, M.T.E.P., VISSER, E.S., DU PLESSIS, J.L., VOGEL, S.W. & ALLSOPP, B.A. 1997. Different organisms associated with heartwater as shown by analysis of 16S ribosomal RNA gene sequences. *Veterinary Parasitology* **71**: 283-300.
- ALLSOPP, M.T.E.P., THERON, J., COETZEE, M.L., DUNSTERVILLE, M.T. & ALLSOPP, B.A. 1999. The occurrence of *Theileria* and *Cowdria* parasites in African buffalo (*Syncerus caffer*) and their associated *Amblyomma hebraeum* ticks. *Onderstepoort Journal of Veterinary Research* **66**: 245-249.
- ALLSOPP, B.A., BEZUIDENHOUT, J.D. & PROZESKY, L. 2004. Heartwater. In: Infectious diseases of livestock 2nd ed. Volume 1. Edited by J.A.W. Coetzer & R.C. Tustin. Oxford: University Press, pp 507-535.
- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

- ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.
- ANDERSSON, S.G.E. & KURLAND, C.G. 1998. Reductive evolution of resident genomes. *Trends in Microbiology* **6**: 263-268.
- ANDERSSON, S.G.E., ZOMORODIPOUR, A., ANDERSSON, J.O., SICHERITZ-PONTÉN, T., ALSMARK, U.C.M., PODOWSKI, R.M., NÄSLUND, A.K., ERIKSSON, A.S., WINKLER, H.H. & KURLAND, C.G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133-140.
- ANSORGE, W.J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology* **25**: 195-203.
- Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.
- ARIEL, N., ZVI, A., MAKAROVA, K.S., CHITLARU, T., ELHANANY, E., VELAN, B., COHEN, S., FRIEDLANDER, A.M. & SHAFFERMAN, A. 2003. Genome-based bioinformatic selection of chromosomal *Bacillus anthracis* putative vaccine candidates coupled with proteomic identification of surface-associated antigens. *Infection and Immunity* **71**: 4563-4579.
- BANKIER, A.T., WESTON, K.M. & BARRELL, B.G. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods in Enzymology* **155**: 51-93.
- BAO, W., KUMAGAI, Y., NIU, H., YAMAGUCHI, M., MIURA, K. & RIKIHISA, Y. 2009. Four VirB6 paralogs and VirB9 are expressed and interact in *Ehrlichia chaffeensis*-containing vacuoles. *Journal of Bacteriology* **191**: 278-286.
- BARBET, A.F., WHITMIRE, W.M., KAMPER, S.M., SIMBI, B.H., GANTA, R.R., MORELAND, A.L., MWANGI, D.M., MCGUIRE, T.C. & MAHAN, S.M. 2001. A subset of *Cowdria ruminantium* genes important for immune recognition and protection. *Gene* **275**: 287-298.
- BARRÉ, N., UILENBERG, G., MOREL, P.C. & CAMUS, E. 1987. Danger of introducing heartwater on to the American mainland: potential role of indigenous and exotic *Amblyomma* ticks. *Onderstepoort Journal of Veterinary Research* **54**: 405-417.
- BASAVANNA, S., KHANDAVILLI, S., YUSTE, J., COHEN, J.M., HOSIE, A.H., WEBB, A.J., THOMAS, G.H. & BROWN, J.S. 2009. Screening of *Streptococcus pneumoniae* ABC transporter

mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infection and Immunity* **77**: 3412-3423.

- BATEMAN, A., COIN, L., DURBIN, R., FINN, R.D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA, A., MARSHALL, M., MOXON, S., SONNHAMMER, E.L., *et al.* 2004. The Pfam protein families database. *Nucleic Acids Research* **32**: D138-D141.
- BEKAL, S., CRAIG, J.P., HUDSON, M.E., NIBLACK, T.L., DOMIER, L.L. & LAMBERT, K.N. 2008. Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Molecular Genetics and Genomics* **279**: 535-543.
- BEKKER, C.P., POSTIGO, M., TAOUFIK, A., BELL-SAKYI, L., FERRAZ, C., MARTINEZ, D. & JONGEJAN, F. 2005. Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. *Journal of Bacteriology* **187**: 4782-4791.
- BELL-SAKYI, L., ZWEYGARTH, E., BLOUIN, E.F., GOULD, E.A. & JONGEJAN, F. 2007. Tick cell lines: tools for tick and tick-borne disease research. *Trends in Parasitology* **23**: 450-457.
- BENNETT, S.T., BARNES, C., COX, A., DAVIES, L. & BROWN, C. 2005. Toward the 1,000 dollars human genome. *Pharmacogenomics* **6**: 373-382.
- BENSON, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573-580.
- BENTLEY, D.R. 2006. Whole-genome re-sequencing. *Current Opinion in Genetics and Development* **16**: 545-552.
- BENZ, I. & SCHMIDT, M.A. 2002. Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology* **45**: 267-276.
- BESEMER, J., LOMSADZE, A. & BORODOVSKY, M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* **29**: 2607-2618.
- BEZUIDENHOUT, J.D., PATERSON, C.L. & BARNARD, B.J.H. 1985. *In vitro* cultivation of *Cowdria ruminantium*. *Onderstepoort Journal of Veterinary Research* **52**: 113-120.
- BEZUIDENHOUT, J.D. 1987. Natural transmission of heartwater. *Onderstepoort Journal of Veterinary Research* **54**: 349-351.

- BONFIELD, J.K., SMITH, K. & STADEN, R. 1995. A new DNA sequence assembly program. *Nucleic Acids Research* **23**: 4992-4999.
- BORK, P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* **17**: 363-374.
- Bovine Genome Sequencing and Analysis Consortium. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**: 522-528.
- BRASLAVSKY, I., HEBERT, B., KARTALOV, E. & QUAKE, S.R. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 3960-3964.
- BRAYTON, K.A., BOTHMA, G.C., VOGEL, S.W. & ALLSOPP, B.A. 1997a. Development of the OPgunTM for bombardment of animal tissues. *Onderstepoort Journal of Veterinary Research* **64**: 153-156.
- BRAYTON, K.A., FEHRSEN, J., DE VILLIERS, E.P., VAN KLEEF, M. & ALLSOPP, B.A. 1997b. Construction and initial analysis of a representative λ ZAPII expression library of the intracellular rickettsia *Cowdria ruminantium*: cloning of *map1* and three other *Cowdria* genes. *Veterinary Parasitology* **72**: 185-199.
- BRAYTON, K.A., DE VILLIERS, E.P., FEHRSEN, J., NXOMANI, C., COLLINS, N.E. & ALLSOPP, B.A. 1999. *Cowdria ruminantium* DNA is unstable in a SuperCos1 library. *Onderstepoort Journal of Veterinary Research* **66**: 111-117.
- BRAYTON, K.A., COLLINS, N.E., VAN STRIJP, F. & ALLSOPP, B.A. 2003. Preparation of *Ehrlichia ruminantium* challenge material for quantifiable and reproducible challenge in mice and sheep. *Veterinary Pathology* **112**: 63-73.
- BRAYTON, K.A., KAPPMAYER, L.S., HERNDON, D.R., DARK, M.J., TIBBALS, D.L., PALMER, G.H., MCGUIRE, T.C. & KNOWLES, D.P. Jr. 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 844-849.
- BRENNAN, R.G. & MATTHEWS, B.W. 1989. The helix-turn-helix DNA binding motif. *Journal of Biological Chemistry* **264**: 1903-1906.
- BROSCH, R., GORDON, S.V., BILLAULT, A., GARNIER, T., EIGLMEIER, K., SORAVITO, C., BARRELL, B.G. & COLE, S.T. 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial

- artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infection and Immunity* **66**: 2221-2229.
- BROWN, W.C., MCELWAIN, T.F., RUEF, B.J., SUAREZ, C.E., SHKAP, V., CHITKO-MCKOWN, C.G., TUO, W., RICE-FICHT, A.C. & PALMER, G.H. 1996. *Babesia bovis* rhostry-associated protein 1 is immunodominant for T helper cells of immune cattle and contains T-cell epitopes conserved among geographically distant *B. bovis* strains. *Infection and Immunity* **64**: 3341-3350.
- BROWN, W.C., MCELWAIN, T.F., PALMER, G.H., CHANTLER, S.E. & ESTES, D.M. 1999. Bovine CD4(+) T-lymphocyte clones specific for rhostry-associated protein 1 of *Babesia bigemina* stimulate enhanced immunoglobulin G1 (IgG1) and IgG2 synthesis. *Infection and Immunity* **67**: 155-164.
- BROWN, J.S., OGUNNIYI, A.D., WOODROW, M.C., HOLDEN, D.W. & PATON, J.C. 2001. Immunization with components of two iron uptake ABC transporters protects mice against systemic *Streptococcus pneumoniae* infection. *Infection and Immunity* **69**: 6702-6706.
- BYROM, B., BARBET, A.F., OBWOLO, M. & MAHAN, S.M. 2000. CD8+ T cell knockout mice are less susceptible to *Cowdria ruminantium* infection than athymic CD4+ T cell knockout, and normal C57BL/6 mice. *Veterinary Parasitology* **93**: 159-172.
- CAMEJO, A., BUCHRIESER, C., COUVÉ, E., CARVALHO, F., REIS, O., FERREIRA, P., SOUSA, S., COSSART, P. & CABANES, D. 2009. *In vivo* transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection. *PLoS Pathogens* **5**: e1000449.
- CAMUS, E. & BARRÉ, N. 1987. Diagnosis of heartwater in the live animal: experiences with goats in Guadeloupe. *Onderstepoort Journal of Veterinary Research* **54**: 291-294.
- CAMUS, E. & BARRÉ, N. 1995. Vector situation of tick-borne diseases in the Caribbean Islands. *Veterinary Parasitology* **57**: 167-176.
- CARVER, T.J., RUTHERFORD, K.M., BERRIMAN, M., RAJANDREAM, M.A., BARRELL, B.G. & PARKHILL, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422-3423.
- CATUREGLI, P., ASANOVICH, K.M., WALLS, J.J., BAKKEN, J.S., MADIGAN, J.E., POPOV, V.L. & DUMLER, J.S. 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infection and Immunity* **68**: 5277-5283.
- C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**: 2012-2018.

- CERRETTI, D.P., DEAN, D., DAVIS, G.R., BEDWELL, D.M. & NOMURA, M. 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. *Nucleic Acids Research* **11**: 2599-2616.
- CHABALGOITY, J.A., BAZ, A., RIAL, A. & GRILLE, S. 2007. The relevance of cytokines for development of protective immunity and rational design of vaccines. *Cytokine & Growth Factor* **18**: 195-207.
- CHAKRAVARTI, D.N., FISKE, M.J., FLETCHER, L.D. & ZAGURSKY, R.J. 2001. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* **19**: 601-612.
- CHAPLIN, P.J., DE ROSE, R., BOYLE, J.S., MCWATERS, P., KELLY, J., TENNENT, J.M., LEW, A.M. & SCHEERLINCK, J.P. 1999. Targeting improves the efficacy of a DNA vaccine against *Corynebacterium pseudotuberculosis* in sheep. *Infection and Immunity* **67**: 6434-6438.
- CHEN, J.M., COOPER, D.N., CHUZHANOVA, N., FÉREC, C. & PATRINOS, G.P. 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8**: 762-775.
- CHOO, K.H., TAN, T.W. & RANGANATHAN, S. 2009. A comprehensive assessment of N-terminal signal peptides prediction methods. *BMC Bioinformatics* **10**: S2.
- CHRISTIE, P.J. 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Molecular Microbiology* **40**: 294-305.
- CITTI, C., KIM, M.F. & WISE, K.S. 1997. Elongated versions of Vlp surface lipoproteins protect *Mycoplasma hyorhinis* escape variants from growth-inhibiting host antibodies. *Infection and Immunity* **65**: 1773-1785.
- COLLINS, N.E., PRETORIUS, A., VAN KLEEF, M., BRAYTON, K.A., ALLSOPP, M.T., ZWEYGARTH, E. & ALLSOPP, B.A. 2003. Development of improved attenuated and nucleic acid vaccines for heartwater. *Developments in Biologicals* **114**: 121-136.
- COLLINS, N.E., LIEBENBERG, J., DE VILLIERS, E.P., BRAYTON, K.A., LOUW, E., PRETORIUS, A., FABER, F.E., VAN HEERDEN, H., JOSEMANS, A., VAN KLEEF, M., *et al.* 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 838-843.

- CORE, L. & PEREGO, M. 2003. TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in *Bacillus subtilis*. *Molecular Microbiology* **49**: 1509-1522.
- COSSART, P. & ARCHAMBAUD, C. 2009. The bacterial pathogen *Listeria monocytogenes*: an emerging model in prokaryotic transcriptomics. *Journal of Biology* **8**: 107.
- COWDRY, E.V. 1925a. Studies on the etiology of heartwater. I. Observation of a *Rickettsia*, *Rickettsia ruminantium* in the tissues of infected animals. *Journal of Experimental Medicine* **42**: 231-252.
- COWDRY, E.V. 1925b. Studies on the etiology of heartwater. II. *Rickettsia ruminantium* in the tissues of ticks transmitting the disease. *Journal of Experimental Medicine* **42**: 253-274.
- COWDRY, E.V. 1926. Studies on the etiology of heartwater. III. The multiplication of *Rickettsia ruminantium* within the endothelial cells of infected animals and their discharge into the circulation. *Journal of Experimental Medicine* **44**: 803-814.
- CROCQUET-VALDES, P.A., MCBRIDE, J.W., YU, X.J. & WALKER, D.H. 2003. Molecular characterization of the 28 kDa multigene locus of *Ehrlichia muris*. *Annals of the New York Academy of Sciences* **990**: 714-716.
- CURASSON, M.G. 1943. *Traité de protozoologie veterinaire et comparée. Trypanosomes, Volume I.* Paris: Vigot Frères, pp 267-279.
- DAS, A. & LJUNGDAHL, L.G. 2003. *Clostridium pasteurianum* F₁F₀ ATP synthase: operon, composition, and some properties. *Journal of Bacteriology* **185**: 5527-5535.
- DAS, A. & PAZOUR, G.J. 1989. Delineation of the regulatory region sequences of *Agrobacterium tumefaciens* virB operon. *Nucleic Acids Research* **17**: 4541-4550.
- D'AURIA, G., JIMÉNEZ, N., PERIS-BONDIA, F., PELAZ, C., LATORRE, A. & MOYA, A. 2008. Virulence factor rtx in *Legionella pneumophila*, evidence suggesting it is a modular multifunctional protein. *BMC Genomics* **9**: 14.
- DAVIS, H.L., MICHEL, M.L., MANCINI, M., SCHLEEF, M. & WHALEN, R.G. 1994. Direct gene transfer in skeletal muscle: plasmid DNA-based immunization against the hepatitis B virus surface antigen. *Vaccine* **12**: 1503-1509.
- DEBOUCK, C. & GOODFELLOW, P.N. 1999. DNA microarrays in drug discovery and development. *Nature Genetics* **21**: 48-50.

- DECKERS-HEBESTREIT, G. & ALTENDORF, K. 1996. The F₀F₁ -type ATP synthases of bacteria: structure and function of the F₀ complex. *Annual Review of Microbiology* **50**: 791-824.
- DE LA FUENTE, J., GARCIA-GARCIA, J.C., BLOUIN, E.F. & KOCAN, K.M. 2003. Characterization of the functional domain of major surface protein 1a involved in adhesion of the rickettsia *Anaplasma marginale* to host cells. *Veterinary Microbiology* **91**: 265-283.
- DE LA FUENTE, J., GARCIA-GARCIA, J.C., BARBET, A.F., BLOUIN, E.F. & KOCAN, K.M. 2004. Adhesion of outer membrane proteins containing tandem repeats of *Anaplasma* and *Ehrlichia* species (Rickettsiales: Anaplasmataceae) to tick cells. *Veterinary Microbiology* **98**: 313-322.
- DE LA FUENTE, J., KOCAN, K.M., BLOUIN, E.F., ZIVKOVIC, Z., NARANJO, V., ALMAZÁN, C., ESTEVES, E., JONGEJAN, F., DAFFRE, S. & MANGOLD, A.J. 2010. Functional genomics and evolution of tick-*Anaplasma* interactions and vaccine development. *Veterinary Parasitology* **167**: 175-186.
- DELCHER, A.L., HARMON, D., KASIF, S., WHITE, O. & SALZBERG, S.L. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* **27**: 4636-4641.
- DEMEREK, M., ADELBERG, E.A., CLARK, A.J. & HARTMAN, P.E. 1966. A proposal for a uniform nomenclature in bacterial genetics. *Genetics* **54**: 61-76.
- DE VILLIERS, E.P., BRAYTON, K.A., ZWEYGARTH, E. & ALLSOPP, B.A. 2000. Genome size and genetic map of *Cowdria ruminantium*. *Microbiology* **146**: 2627-2634.
- DIGUISTINI, S., LIAO, N.Y., PLATT, D., ROBERTSON, G., SEIDEL, M., CHAN, S.K., DOCKING, T.R., BIROL, I., HOLT, R.A., HIRST, M., *et al.* 2009. *De novo* genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology* **10**: R94.
- DIXON, R.W. 1898. Heartwater experiments. *Agricultural Journal of the Cape of Good Hope* **12**: 754-760.
- DOHM, J.C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2007. SHARCGS, a fast and highly accurate short-read assembly algorithm for *de novo* genomic sequencing. *Genome Research* **17**: 1697-1706.
- DOOLITTLE, R.F. 1981. Similar amino acid sequences: chance or common ancestry? *Science* **214**: 149-159.

- DORIA-ROSE, N.A. & HAIGWOOD, N.L. 2003. DNA vaccine strategies: candidates for immune modulation and immunization regimens. *Methods* **31**: 207-216.
- DORO, F., LIBERATORI, S., RODRÍGUEZ-ORTEGA, M.J., RINAUDO, C.D., ROSINI, R., MORA, M., SCARSELLI, M., ALTINDIS, E., D'AURIZIO, R., STELLA, M., *et al.* 2009. Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for Group B *Streptococcus* hypervirulent strain COH1. *Molecular & Cellular Proteomics* **8**: 1728-1737.
- DOYLE, C.K., ZHANG, X., POPOV, V.L. & MCBRIDE, J.W. 2005. An immunoreactive 38-kilodalton protein of *Ehrlichia canis* shares structural homology and iron-binding capacity with the ferric ion-binding protein family. *Infection and Immunity* **73**: 62-69.
- DREW, D.R., BOYLE, J.S., LEW, A.M., LIGHTOWLERS, M.W., CHAPLIN, P.J. & STRUGNELL, R.A. 2001. The comparative efficacy of CTLA-4 and L-selectin targeted DNA vaccines in mice and sheep. *Vaccine* **19**: 4417-4428.
- DROEGE, M. & HILL, B. 2008. The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* **136**: 3-10.
- DUMLER, J.S., BARBET, A.F., BEKKER, C.P.J., DASCH, G.A., PALMER, G.H., RAY, S.C., RIKIHISA, Y. & RURANGIRWA, F.R. 2001. Reorganisation of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* and *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and HE agent as subjective synonyms of *Ehrlichia phagocytophila*. *International Journal of Systematic and Evolutionary Microbiology* **51**: 2145-2165.
- DUNHAM, S.P. 2002. The application of nucleic acid vaccines in veterinary medicine. *Research in Veterinary Science* **73**: 9-16.
- DU PLESSIS, J.L. & KÜMM, N.A.L. 1971. The passage of *Cowdria ruminantium* in mice. *Journal of the South African Veterinary Medical Association* **42**: 217-221.
- DU PLESSIS, J.L. 1985. A method for determining the *Cowdria ruminantium* infection rate of *Amblyomma hebraeum*: effects in mice injected with tick homogenates. *Onderstepoort Journal of Veterinary Research* **52**: 55-61.
- DU PLESSIS, J.L. & MALAN, L. 1987. The non-specific resistance of cattle to heartwater. *Onderstepoort Journal of Veterinary Research* **54**: 333-336.

- DU PLESSIS, J.L., VAN GAS, L., OLIVIER, J.A. & BEZUIDENHOUT, J.D. 1989. The heterogeneity of *Cowdria ruminantium* isolates: cross immunity and serology in sheep and pathogenicity to mice. *Onderstepoort Journal of Veterinary Research* **56**: 195-201.
- DU PLESSIS, J.L., BERCHE, P. & VAN GAS, L. 1991. T cell-mediated immunity to *Cowdria ruminantium* in mice: the protective role of Lyt-2+ T cells. *Onderstepoort Journal of Veterinary Research* **58**: 171-179.
- DU PLESSIS, J.L., BEZUIDENHOUT, J.D., BRETT, M.S., CAMUS, E., JONGEJAN, F., MAHAN, S.M. & MARTINEZ, D. 1993. The sero-diagnosis of heartwater: a comparison of five tests. *Revue d'Élevage et de Médecine Vétérinaire des Pays Tropicaux* **46**: 123-129.
- EDINGTON, A. 1898. Heartwater. *Agricultural Journal of the Cape of Good Hope* **12**: 748-754.
- EDWARDS, A., VOSS, H., RICE, P., CIVITELLO, A., STEGEMANN, J., SCHWAGER, C., ZIMMERMANN, J., ERFLE, H., CASKEY, C.T. & ANSORGE, W. 1990. Automated DNA sequencing of the human HPRT locus. *Genomics* **6**: 593-608.
- ESTEVEZ, I., MARTINEZ, D. & TOTTE, P. 2004. Identification of *Ehrlichia ruminantium* (Gardel strain) IFN- γ inducing proteins after vaccination with a killed vaccine. *Veterinary Microbiology* **100**: 233-240.
- EWING, B., HILLIER, L., WENDL, M.C. & GREEN, P. 1998. Base-calling automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**: 175-185.
- FELEK, S., HUANG, H. & RIKIHISA, Y. 2003. Sequence and expression analysis of *virB9* of the type IV secretion system of *Ehrlichia canis* strains in ticks, dogs, and cultured cells. *Infection and Immunity* **71**: 6063-6067.
- FELSHEIM, R.F., HERRON, M.J., NELSON, C.M., BURKHARDT, N.Y., BARBET, A.F., KURTTI, T.J. & MUNDERLOH, U.G. 2006. Transformation of *Anaplasma phagocytophilum*. *BMC Biotechnology* **6**: 42.
- FELSHEIM, R.F., CHÁVEZ, A.S., PALMER, G.H., CROSBY, L., BARBET, A.F., KURTTI, T.J. & MUNDERLOH, U.G. 2010. Transformation of *Anaplasma marginale*. *Veterinary Parasitology* **167**: 167-174.
- FENN, K. & BLAXTER, M. 2006. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends in Parasitology* **22**: 60-65.

- FINLAY, B.B. & FALKOW, S. 1997. Common themes in microbial pathogenicity revisited. *Microbiology and Molecular Biology Reviews* **61**: 136-169.
- FLACH, E.J., WOODFORD, J.D., MORZARIA, S.P., DOLAN, T.T. & SHAMBWANA, I. 1990. Identification of *Babesia bovis* and *Cowdria ruminantium* on the island of Unguja, Zanzibar. *The Veterinary Record* **126**: 57-59.
- FLEISCHMANN, R.D., ADAMS, M.D., WHITE, O., CLAYTON, R.A., KIRKNESS, E.F., KERLAVAGE, A.R., BULT, C.J., TOMB, J.F., DOUGHERTY, B.A., MERRICK, J.M., *et al.* 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.
- FOSTER, J., GANATRA, M., KAMAL, I., WARE, J., MAKAROVA, K., IVANOVA, N., BHATTACHARYYA, A., KAPATRAL, V., KUMAR, S., POSFAI, J., *et al.* 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biology* **3**: e121.
- FRANGEUL, L., NELSON, K.E., BUCHRIESER, C., DANCHIN, A., GLASER, P. & KUNST, F. 1999. Cloning and assembly strategies in microbial genome projects. *Microbiology* **145**: 2625-2634.
- FREDRICKS, D.N. 2006. Introduction to the Rickettsiales and other intracellular prokaryotes. In: The prokaryotes: A handbook on the biology of bacteria, 3rd ed. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer and E. Stackebrandt. New York: Springer, pp 457-466.
- FRISHMAN, D., MIRONOV, A., MEWES, H.W. & GELFAND, M. 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Research* **26**: 2941-2947.
- FRUTOS, R., VIARI, A., FERRAZ, C., MORGAT, A., EYCHENIÉ, S., KANDASSAMY, Y., CHANTAL, I., BENSAID, A., COISSAC, E., VACHIERY, N., *et al.* 2006. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *Journal of Bacteriology* **188**: 2533-2542.
- FRUTOS, R., VIARI, A., VACHIERY, N., BOYER, F. & MARTINEZ, D. 2007. *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends in Parasitology* **23**: 414-419.
- FU, X., FU, N., GUO, S., YAN, Z., XU, Y., HU, H., MENZEL, C., CHEN, W., LI, Y., ZENG, R. & KHAITOVICH, P. 2009. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**: 161.

- GALINDO, R.C., AYOUBI, P., GARCÍA-PÉREZ, A.L., NARANJO, V., KOCAN, K.M., GORTAZAR, C. & DE LA FUENTE, J. 2008. Differential expression of inflammatory and immune response genes in sheep infected with *Anaplasma phagocytophilum*. *Veterinary Immunology and Immunopathology* **126**: 27-34.
- GARCIA-GARCIA, J.C., DE LA FUENTE, J., BELL-EUNICE, G., BLOUIN, E.F. & KOCAN, K.M. 2004. Glycosylation of *Anaplasma marginale* major surface protein 1a and its putative role in adhesion to tick cells. *Infection and Immunity* **72**: 3022-3030.
- GARDNER, M.J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R.W., CARLTON, J.M., PAIN, A., NELSON, K.E., BOWMAN, S., *et al.* 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498-511.
- GARDY, J.L., LAIRD, M.R., CHEN, F., REY, S., WALSH, C.J., ESTER, M. & BRINKMAN, F.S.L. 2005. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localisation and insights gained from comparative proteome analysis. *Bioinformatics* **21**: 617-623.
- GIOVANNONI, S.J., TRIPP, H.J., GIVAN, S., PODAR, M., VERGIN, K.L., BAPTISTA, D., BIBBS, L., EADS, J., RICHARDSON, T.H., NOORDEWIER, M., *et al.* 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242-1245.
- GIULIANI, M.M., ADU-BOBIE, J., COMANDUCCI, M., ARICÒ, B., SAVINO, S., SANTINI, L., BRUNELLI, B., BAMBINI, S., BIOLCHI, A., CAPECCHI, B., *et al.* 2006. A universal vaccine for serogroup B meningococcus. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 10834-10839.
- GOLDBERG, S.M., JOHNSON, J., BUSAM, D., FELDBLYUM, T., FERRIERA, S., FRIEDMAN, R., HALPERN, A., KHOURI, H., KRAVITZ, S.A., LAURO, F.M., *et al.* 2006. A Sanger/ pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11240-11245.
- GRANDI, G. 2001. Antibacterial vaccine design using genomics and proteomics. *Trends in Biotechnology* **19**: 181-188.
- GRANDI, G. 2003. Rational antibacterial vaccine design through genomic technologies. *International Journal for Parasitology* **33**: 615-620.
- GRIFANTINI, R., BARTOLINI, E., MUZZI, A., DRAGHI, M., FRIGIMELICA, E., BERGER, J., RATTI, G., PETRACCA, R., GALLI, G., AGNUSDEI, M., *et al.* 2002. Previously unrecognized vaccine

- candidates against group B meningococcus identified by DNA microarrays. *Nature Biotechnology* **20**: 914-921.
- GUPTA, R.S. & MOK, A. 2007. Phylogenomics and signature proteins for the alpha Proteobacteria and its main groups. *BMC Microbiology* **7**: 106.
- HAIG, D.A., ALEXANDER, R.A. & WEISS, K.E. 1954. Treatment of heartwater with tetramycin. *Journal of the South African Veterinary Medical Association* **25**: 45-48.
- HALL, N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *The Journal of Experimental Biology* **209**: 1518-1525.
- HARDING, C.V., RAMACHANDRA, L. & WICK, M.J. 2003. Interaction of bacteria with antigen presenting cells: influences on antigen presentation and antibacterial immunity. *Current Opinion in Immunology* **15**: 112-119.
- HARRIS, T.D., BUZBY, P.R., BABCOCK, H., BEER, E., BOWERS, J., BRASLAVSKY, I., CAUSEY, M., COLONELL, J., DIMEO, J., EFCAVITCH, J.W., *et al.* 2008. Single-molecule DNA sequencing of a viral genome. *Science* **320**: 106-109.
- HARRISON, R.G. 2000. Expression of soluble heterologous proteins via fusion with NusA protein. *inNovations* **11**: 4-7.
- HATCH, T.P. 1999. Developmental biology. In: *Chlamydia: intracellular biology, pathogenesis, and immunity*. Edited by R.S. Stephens. Washington, D.C.: American Society for Microbiology, pp 29-67.
- HICKLING, J.K. 1998. Measuring human T-lymphocyte function. *Expert Reviews in Molecular Medicine* **1998**: 1-20.
- HIGGINS, C.F. 2001. ABC transporters: physiology, structure and mechanism – an overview. *Research in Microbiology* **152**: 205-210.
- HILLIER, L.W., MARTH, G.T., QUINLAN, A.R., DOOLING, D., FEWELL, G., BARNETT, D., FOX, P., GLASSCOCK, J.I., HICKENBOTHAM, M., HUANG, W., *et al.* 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* **5**: 183-188.
- HOLT, K.E., PARKHILL, J., MAZZONI, C.J., ROUMAGNAC, P., WEILL, F.X., GOODHEAD, I., RANCE, R., BAKER, S., MASKELL, D.J., WAIN, J., *et al.* 2008. High-throughput sequencing

- provides insights into genome variation and evolution in *Salmonella* Typhi. *Nature Genetics* **40**: 987-993.
- HOOD, D.W., DEADMAN, M.E., JENNINGS, M.P., BISERCIC, M., FLEISCHMANN, R.D., VENTER, J.C. & MOXON, E.R. 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 11121-11125.
- HOTOPP, J.C., LIN, M., MADUPU, R., CRABTREE, J., ANGIUOLI, S.V., EISEN, J., SESHADRI, R., REN, Q., WU, M., UTTERBACK, T.R. *et al.* 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genetics* **2**: e21.
- HUGHES, D. 2000a. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biology* **1**: review 0006.1-0006.8.
- HUGHES, D. 2000b. Co-evolution of the *tuf* genes links gene conversion with the generation of chromosomal inversions. *Journal of Molecular Biology* **297**: 355-364.
- HUTCHEON, D. 1900. History of heartwater. *Agricultural Journal of the Cape of Good Hope* **17**: 410-417.
- HUYGEN, K. 2003. On the use of DNA vaccines for the prophylaxis of mycobacterial diseases. *Infection and Immunity* **71**: 1613-1621.
- IBBA, M., CURNOW, A.W. & SOLL, D. 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. *Trends in Biochemical Sciences* **22**: 39-42.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- IOANNIDIS, P., HOTOPP, J.C., SAPOUNTZIS, P., SIOZIOS, S., TSIAMIS, G., BORDENSTEIN, S.R., BALDO, L., WERREN, J.H. & BOURTZIS, K. 2007. New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics* **8**: 182.
- JIANG, G., CHAROENVIT, Y., MORENO, A., BARACEROS, M.F., BANANIA, G., RICHIE, N., ABOT, S., GANESHAN, H., FALLARME, V., PATTERSON, N.B., *et al.* 2007. Induction of multi-antigen multi-stage immune responses against *Plasmodium falciparum* in rhesus monkeys, in the absence of antigen interference, with heterologous DNA prime/poxvirus boost immunization. *Malaria Journal* **6**: 135.

- JONGEJAN, F. 1991. Protective immunity to heartwater (*Cowdria ruminantium* infection) is acquired after vaccination with in vitro-attenuated rickettsiae. *Infection and Immunity* **59**: 729-731.
- JOSEMANS, A.I. & ZWEYGARTH, E. 2002. Amino acid content of cell cultures infected with *Cowdria ruminantium* propagated in a protein-free medium. *Annals of the New York Academy of Sciences* **969**: 141-146.
- JUHAS, M., CROOK, D.W. & HOOD, D.W. 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cellular Microbiology* **10**: 2377-2386.
- JURINKE, C., VAN DEN BOOM, D., CANTOR, C.R. & KOSTER, H. 2002. The use of MassARRAY technology for high throughput genotyping. *Advances in Biochemical Engineering / Biotechnology* **77**: 57-74.
- KÄLL, L., KROGH, A. & SONNHAMMER, E.L.L. 2004. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**: 1027-1036.
- KALMAN, S., MITCHELL, W., MARATHE, R., LAMMEL, C., FAN, J., HYMAN, R.W., OLINGER, L., GRIMWOOD, J., DAVIS, R.W. & STEPHENS, R.S. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature Genetics* **21**: 385-389.
- KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**: 27-30.
- KANEKO, T., NAKAMURA, Y., SATO, S., ASAMIZU, E., KATO, T., SASAMOTO, S., WATANABE, A., IDESAWA, K., ISHIKAWA, A., KAWASHIMA, K., *et al.* 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* **7**: 331-338.
- KANYARI, P.W. & KAGIRA, J. 2000. The role of parasitic diseases as causes of mortality in cattle in a high potential area of central Kenya: a quantitative analysis. *Onderstepoort Journal of Veterinary Research* **67**: 157-161.
- KASIANOWICZ, J.J., BRANDIN, E., BRANTON, D. & DEAMER, D.W. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 13770-13773.
- KATHAPERUMAL, K., KUMANAN, V., MCDONOUGH, S., CHEN, L.H., PARK, S.U., MOREIRA, M.A., AKEY, B., HUNTLEY, J., CHANG, C.F. & CHANG, Y.F. 2009. Evaluation of immune responses and protective efficacy in a goat model following immunization with a cocktail of

- recombinant antigens and a polyprotein of *Mycobacterium avium* subsp. *paratuberculosis*. *Vaccine* **27**: 123-135.
- KAUSHIK, D.K. & SEHGAL, D. 2008. Developing antibacterial vaccines in genomics and proteomics era. *Scandinavian Journal of Immunology* **67**: 544-552.
- KENNEDY, N.J., SPITHILL, T.W., TENNENT, J., WOOD, P.R. & PIEDRAFITA, D. 2006. DNA vaccines in sheep: CTLA-4 mediated targeting and CpG motifs enhance immunogenicity in a DNA prime/protein boost strategy. *Vaccine* **24**: 970-979.
- KOCAN, K.M. & BEZUIDENHOUT, J.D. 1987. Morphology and development of *Cowdria ruminantium* in *Amblyomma* ticks. *Onderstepoort Journal of Veterinary Research* **54**: 177-182.
- KOCAN, K.M., BEZUIDENHOUT, J.D. & HART, A. 1987. Ultrastructural features of *Cowdria ruminantium* in midgut epithelial cells and salivary glands of nymphal *Amblyomma hebraeum*. *Onderstepoort Journal of Veterinary Research* **54**: 87-92.
- KOCK, N.D., VAN VLIET, A.H.M., CHARLTON, K. & JONGEJAN, F. 1995. Detection of *Cowdria ruminantium* in blood and bone marrow samples from clinically normal, free-ranging Zimbabwean wild ungulates. *Journal of Clinical Microbiology* **33**: 2501-2504.
- KODAMA, K., KAWAMURA, S., YASUKAWA, M. & KOBAYASHI, Y. 1987. Establishment and characterization of a T-cell line specific for *Rickettsia tsutsugamushi*. *Infection and Immunity* **55**: 2490-2495.
- KOIDE, T., ZAINI, P.A., MOREIRA, L.M., VÊNICIO, R.Z., MATSUKUMA, A.Y., DURHAM, A.M., TEIXEIRA, D.C., EL-DORRY, H., MONTEIRO, P.B., DA SILVA, A.C., *et al.* 2004. DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. *Journal of Bacteriology* **186**: 5442-5449.
- KOLPAKOV, R., BANA, G. & KUCHEROV, G. 2003. mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research* **31**: 3672-3678.
- KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E.L.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.
- KUMAGAI, Y., HUANG, H. & RIKIHISA, Y. 2008. Expression and porin activity of P28 and OMP-1F during intracellular *Ehrlichia chaffeensis* development. *Journal of Bacteriology* **190**: 3597-3605.

- KUNST, F., OGASAWARA, N., MOSZER, I., ALBERTINI, A.M., ALLONI, G., AZEVEDO, V., BERTERO, M.G., BESSIÈRES, P., BOLOTIN, A., BORCHERT, S., *et al.* 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256.
- LANGDON, S.P. 2004. Cell culture contamination: an overview. *Methods in Molecular Medicine* **88**: 309-317.
- LAWSON, M.J., JIAO, J., FAN, W. & ZHANG, L. 2009. A pattern analysis of gene conversion literature. *Comparative and Functional Genomics* **2009**: 761512.
- LEUNG, W.H., MENG, Z.Q., HUI, G. & HO, W.K. 2004. Expression of an immunologically reactive merozoite surface protein (MSP-1₄₂) in *E. coli*. *Biochimica et Biophysica Acta* **1675**: 62-70.
- LEVINSON, G. & GUTMAN, G.A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**: 203-221.
- LIN, J., HUANG, S. & ZHANG, Q. 2002. Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes and Infection* **4**: 325-331.
- LIN, M. & RIKIHISHA, Y. 2003. *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infection and Immunity* **71**: 5324-5331.
- LIN, M., DEN DULK-RAS, A., HOOYKAAS, P.J. & RIKIHISA, Y. 2007. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cellular Microbiology* **9**: 2644-2657.
- LIN, M., ZHANG, C., GIBSON, K. & RIKIHISA, Y. 2009. Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever. *Nucleic Acids Research* **37**: 6076-6091.
- LOBRY, J.R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution* **13**: 660-665.
- LOGAN, L.L., WHYARD, T.C., QUINTERO, J.C. & MEBUS, C.A. 1987. The development of *Cowdria ruminantium* in neutrophils. *Onderstepoort Journal of Veterinary Research* **54**: 197-204.
- LONG, S.W., WHITWORTH, T.J., WALKER, D.H. & YU, X.J. 2005. Overcoming barriers to the transformation of the genus *Ehrlichia*. *Annals of the New York Academy of Sciences* **1063**: 403-410.

- LOPEZ, J.E., SIEMS, W.F., PALMER, G.H., BRAYTON, K.A., MCGUIRE, T.C., NORIMINE, J. & BROWN, W.C. 2005. Identification of novel antigenic proteins in a complex *Anaplasma marginale* outer membrane immunogen by mass spectrometry and genomic mapping. *Infection and Immunity* **73**: 8109-8118.
- LOPEZ, J.E., PALMER, G.H., BRAYTON, K.A., DARK, M.J., LEACH, S.E. & BROWN, W.C. 2007. Immunogenicity of *Anaplasma marginale* type IV secretion system proteins in a protective outer membrane vaccine. *Infection and Immunity* **75**: 2333-2342.
- LORENZEN, N. & LAPATRA, S.E. 2005. DNA vaccines for aquacultured fish. *Revue Scientifique et Technique* **24**: 201-213.
- LOUNSBURY, C.P. 1900. Tick-heartwater experiment. *Agricultural Journal of the Cape of Good Hope* **16**: 682-687.
- LOUW, E., BRAYTON, K.A., COLLINS, N.E., PRETORIUS, A., VAN STRIJP, F. & ALLSOPP, B.A. 2002. Sequencing of a 15-kb *Ehrlichia ruminantium* clone and evaluation of the *cpg1* open reading frame for protection against heartwater. *Annals of the New York Academy of Sciences* **969**: 147-150.
- LOWE, T.M. & EDDY, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955-964.
- LUDWIG, W. & KLENK, H.P. 2001. Overview: A phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Bergey's manual of systematic bacteriology, 2nd ed. Edited by D.R. Boone and R.W. Castenholz. Berlin: Springer-Verlag, pp 49-65.
- LUO, D., NI, B., LI, P., SHI, W., ZHANG, S., HAN, Y., MAO, L., HE, Y., WU, Y. & WANG, X. 2006. Protective immunity elicited by a divalent DNA vaccine encoding both the L7/L12 and Omp16 genes of *Brucella abortus* in BALB/c mice. *Infection and Immunity* **74**: 2734-2741.
- LUO, T., ZHANG, X., WAKEEL, A., POPOV, V.L. & MCBRIDE, J.W. 2008. A variable-length PCR target protein of *Ehrlichia chaffeensis* contains major species-specific antibody epitopes in acidic serine-rich tandem repeats. *Infection and Immunity* **76**: 1572-1580.
- LUO, T., ZHANG, X. & MCBRIDE, J.W. 2009. Major species-specific antibody epitopes of the *Ehrlichia chaffeensis* p120 and *E. canis* p140 orthologs in surface-exposed tandem repeat regions. *Clinical and Vaccine Immunology* **16**: 982-990.
- MADABHUSHI, R.S. 1998. Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. *Electrophoresis* **19**: 224-230.

- MAHAN, S.M., ANDREW, H.R., TEBELE, N., BURRIDGE, M.J. & BARBET, A.F. 1995. Immunization of sheep against heartwater with inactivated *Cowdria ruminantium*. *Research in Veterinary Science* **58**: 46-49.
- MAHAN, S.M., PETER, T.F., SIMBI, B.H., KOCAN, K., CAMUS, E., BARBET, A.F. & BURRIDGE, M.J. 2000. Comparison of efficacy of American and African *Amblyomma* ticks as vectors of heartwater (*Cowdria ruminantium*) infection by molecular analyses and transmission trials. *Journal of Parasitology* **86**: 44-49.
- MAHAN, S.M., SMITH, G.E., KUMBULA, D., BURRIDGE, M.J. & BARBET, A.F. 2001. Reduction in mortality from heartwater in cattle, sheep and goats exposed to field challenge using an inactivated vaccine. *Veterinary Parasitology* **97**: 295-308.
- MAILLARD, J.C. & MAILLARD, N. 1998. Historique du peuplement bovin et de l'introduction de la tique *Amblyomma variegatum* dans les îles françaises des Antilles: Synthèse bibliographique. *Ethnozootechnie* **1**: 19-36.
- MAIONE, D., MARGARIT, I., RINAUDO, C.D., MASIGNANI, V., MORA, M., SCARSELLI, M., TETTELIN, H., BRETTONI, C., IACOBINI, E.T., ROSINI, R., *et al.* 2005. Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen. *Science* **309**: 148-150.
- MANNING, S.D., MOTIWALA, A.S., SPRINGMAN, A.C., QI, W., LACHER, D.W., OUELLETTE, L.M., MLADONICKY, J.M., SOMSEL, P., RUDRIK, J.T., DIETRICH, S.E., *et al.* 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 4868-4873.
- MARGUERAT, S. & BÄHLER, J. 2009. RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* **67**: 569-79.
- MARGULIES, M., EGHOLM, M., ALTMAN, W.E., ATTIYA, S., BADER, J.S., BEMBEN, L.A., BERKA, J., BRAVERMAN, M.S., CHEN, Y.J., CHEN, Z., *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- MARIOTTI, E., MIRABELLI, P., DI NOTO, R., FORTUNATO, G. & SALVATORE, F. 2008. Rapid detection of mycoplasma in continuous cell lines using a selective biochemical test. *Leukemia Research* **32**: 323-326.
- MARTINEZ, D., MAILLARD, J.C., COISNE, S., SHEIKBOUDOU, C. & BENSALD, A. 1994. Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. *Veterinary Immunology and Immunopathology* **41**: 153-163.

- MAVROMATIS, K., DOYLE, C.K., LYKIDIS, A., IVANOVA, N., FRANCINO, M.P., CHAIN, P., SHIN, M., MALFATTI, S., LARIMER, F., COPELAND, A., *et al.* 2006. The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *Journal of Bacteriology* **188**: 4015-4023.
- MAXAM, A.M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 560-564.
- MCBRIDE, J.W., YU, X.J. & WALKER, D.H. 2000. Glycosylation of homologous immunodominant proteins of *Ehrlichia chaffeensis* and *Ehrlichia canis*. *Infection and Immunity* **68**: 13-18.
- MEEUS, P.F. & BARBET, A.F. 2001. Ingenious gene generation. *Trends in Microbiology* **9**: 353-355.
- MEEUS, P.F., BRAYTON, K.A., PALMER, G.H. & BARBET, A.F. 2003. Conservation of a gene conversion mechanism in two distantly related paralogues of *Anaplasma marginale*. *Molecular Microbiology* **47**: 633-643.
- MERRELL, D.S., BUTLER, S.M., QADRI, F., DOLGANOV, N.A., ALAM, A., COHEN, M.B., CALDERWOOD, S.B., SCHOOLNIK, G.K. & CAMILLI, A. 2002. Host induced epidemic spread of the cholera bacterium. *Nature* **417**: 642-645.
- MIN, C.K., YANG, J.S., KIM, S., CHOI, M.S., KIM, I.S. & CHO, N.H. 2008. Genome-based construction of the metabolic pathways of *Orientia tsutsugamushi* and comparative analysis within the Rickettsiales order. *Comparative and Functional Genomics* **2008**: 623145.
- MIYOSHI, S. & SHINODA, S. 2000. Microbial metalloproteases and pathogenesis. *Microbes and Infection* **2**: 91-98.
- MONTIGIANI, S., FALUGI, F., SCARSELLI, M., FINCO, O., PETRACCA, R., GALLI, G., MARIANI, M., MANETTI, R., AGNUSDEI, M., CEVENINI, R., *et al.* 2002. Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. *Infection and Immunity* **70**: 368-379.
- MORTAZAVI, A., WILLIAMS, B.A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621-628.
- MOSHKOVSKI, S.D. 1947. Comments by readers. *Science* **106**: 62.
- MOSMANN, T.R. & SAD, S. 1996. The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunology Today* **17**: 138-146.

- MOXON, E.R., HOOD, D.W., SAUNDERS, N.J., SCHWEDA, E.K. & RICHARDS, J.C. 2002. Functional genomics of pathogenic bacteria. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**: 109-116.
- MULLER KOBOLD, A., MARTINEZ, D., CAMUS, E. & JONGEJAN, F. 1992. Distribution of heartwater in the Caribbean determined on the basis of detection of antibodies to the conserved 32-kilodalton protein of *Cowdria ruminantium*. *Journal of Clinical Microbiology* **30**: 1870-1873.
- MUZZI, A., MASIGNANI, V. & RAPPUOLI, R. 2007. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discovery Today* **12**: 429-439.
- MWANGI, D.M., MAHAN, S.M., NYANJUI, J.K., TARACHA, E.L.N. & MCKEEVER, D.J. 1998. Immunization of cattle by infection with *Cowdria ruminantium* elicits T lymphocytes that recognise autologous infected endothelial cells and monocytes. *Infection and Immunity* **66**: 1855-1860.
- MWANGI, D.M., MCKEEVER, D.J., NYANJUI, J.K., BARBET, A.F. & MAHAN, S.M. 2002. Immunisation of cattle against heartwater by infection with *Cowdria ruminantium* elicits T lymphocytes that recognise major antigenic proteins 1 and 2 of the agent. *Veterinary Immunology and Immunopathology* **85**: 23-32.
- MYERS, G.S., PARKER, D., AL-HASANI, K., KENNAN, R.M., SEEMANN, T., REN, Q., BADGER, J.H., SELENGUT, J.D., DEBOY, R.T., TETTELIN, H., *et al.* 2007. Genome sequence and identification of candidate vaccine antigens from the animal pathogen *Dichelobacter nodosus*. *Nature Biotechnology* **25**: 569-575.
- NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344-1349.
- NAKAMURA, Y., LEPPERT, M., O'CONNELL, P., WOLFF, R., HOLM, T., CULVER, M., MARTIN, C., FUJIMOTO, E., HOFF, M., KUMLIN, E., *et al.* 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**: 1616-1622.
- NAVARRE, W.W. & SCHNEEWIND, O. 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews* **63**: 174-229.
- NEITZ, W.O. 1940. Uleron in the treatment of heartwater. *Journal of the South African Veterinary Medical Association* **11**: 15.

- NEITZ, W.O. & ALEXANDER, R.A. 1941. The immunization of calves against heartwater. *Journal of the South African Veterinary Medical Association* **12**: 103-111.
- NEITZ, W.O. & ALEXANDER, R.A. 1945. Immunization of cattle against heartwater and the control of the tick-borne diseases, redwater, gallsickness and heartwater. *Onderstepoort Journal of Veterinary Science and Animal Industry* **20**: 137-158.
- NEITZ, W.O. 1964. Tick-borne diseases as a hazard in the rearing of calves in South Africa. *Bulletin - Office international des épizooties* **62**: 607-625.
- NEITZ, W.O. 1968. Heartwater. *Bulletin - Office international des épizooties* **70**: 329-336.
- NICHOLS, W.W., LEDWITH, B.J., MANAM, S.V. & TROILO, P.J. 1995. Potential DNA vaccine integration into host cell genome. *Annals of the New York Academy of Sciences* **772**: 30-39.
- NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & VON HEIJNE, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1-6.
- NIEMANN, H.H., SCHUBERT, W.D. & HEINZ, D.W. 2004. Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes and Infection* **6**: 101-112.
- NOVAES, E., DROST, D.R., FARMERIE, W.G., PAPPAS, G.J. Jr., GRATTAPAGLIA, D., SEDEROFF, R.R. & KIRST, M. 2008. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* **9**: 312.
- NYIKA, A., MAHAN, S.M., BURRIDGE, M.J., MCGUIRE, T.C., RURANGIRWA, F. & BARBET, A.F. 1998. A DNA vaccine protects mice against the rickettsial agent *Cowdria ruminantium*. *Parasite Immunology* **20**: 111-119.
- NYIKA, A., BARBET, A.F., BURRIDGE, M.J. & MAHAN, S.M. 2002. DNA vaccination with *map1* gene followed by protein boost augments protection against challenge with *Cowdria ruminantium*, the agent of heartwater. *Vaccine* **20**: 1215-1225.
- OBEREM, P.T. & BEZUIDENHOUT, J.D. 1987. The production of heartwater vaccine. *Onderstepoort Journal of Veterinary Research* **54**: 485-488.
- OGASAWARA, N. & YOSHIKAWA, H. 1992. Genes and their organization in the replication origin region of the bacterial chromosome. *Molecular Microbiology* **6**: 629-634.

- OGATA, H., GOTO, S., SATO, K., FUJIBUCHI, W., BONO, H. & KANEHISA, M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **27**: 29-34.
- OGATA, H., AUDIC, S., BARBE, V., ARTIGUENAVE, F., FOURNIER, P.E., RAOULT, D. & CLAVERIE, J.M. 2000. Selfish DNA in protein-coding genes of *Rickettsia*. *Science* **290**: 347-350.
- OGATA, H., RENESTO, P., AUDIC, S., ROBERT, C., BLANC, G., FOURNIER, P.E., PARINELLO, H., CLAVERIE, J.M. & RAOULT, D. 2005. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biology* **3**: e248.
- OGATA, H., LA SCOLA, B., AUDIC, S., RENESTO, P., BLANC, G., ROBERT, C., FOURNIER, P.E., CLAVERIE, J.M. & RAOULT, D. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genetics* **2**: e76.
- OHASHI, N., UNVER, A., ZHI, N. & RIKIHISA, Y. 1998a. Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of *Ehrlichia canis* and application of the recombinant protein for serodiagnosis. *Journal of Clinical Microbiology* **36**: 2671-2680.
- OHASHI, N., ZHI, N., ZHANG, Y. & RIKIHISA, Y. 1998b. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. *Infection and Immunity* **66**: 132-139.
- OJCIUS, D.M., GACHELIN, G. & DANTRY-VARSAT, A. 1996. Presentation of antigens derived from microorganisms residing in host-cell vacuoles. *Trends in Microbiology* **4**: 53-59.
- OÑATE, A.A., DONOSO, G., MORAGA-CID, G., FOLCH, H., CÉSPEDES, S. & ANDREWS, E. 2005. An RNA vaccine based on recombinant Semliki Forest virus particles expressing the Cu,Zn superoxide dismutase protein of *Brucella abortus* induces protective immunity in BALB/c mice. *Infection and Immunity* **73**: 3294-3300.
- OSADA, N. & INNAN, H. 2008. Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLoS Genetics* **4**: e1000305.
- OZSOLAK, F., PLATT, A.R., JONES, D.R., REIFENBERGER, J.G., SASS, L.E., MCINERNEY, P., THOMPSON, J.F., BOWERS, J., JAROSZ, M. & MILOS, P.M. 2009. Direct RNA sequencing. *Nature* **461**: 814-818.
- PAN, W., RAVOT, E., TOLLE, R., FRANK, R., MOSBACH, R., TÜRBAČHOVA, I. & BUJARD, H. 1999. Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide

- enables synthesis and isolation of full-length protein from *Escherichia coli* and mammalian cells. *Nucleic Acids Research* **27**: 1094-1103.
- PARK, S.U., KATHAPERUMAL, K., MCDONOUGH, S., AKEY, B., HUNTLEY, J., BANNANTINE, J.P. & CHANG, Y.F. 2008. Immunization with a DNA vaccine cocktail induces a Th1 response and protects mice against *Mycobacterium avium* subsp. *paratuberculosis* challenge. *Vaccine* **26**: 4329-4337.
- PARKHILL, J., WREN, B.W., MUNGALL, K., KETLEY, J.M., CHURCHER, C., BASHAM, D., CHILLINGWORTH, T., DAVIES, R.M., FELTWELL, T., HOLROYD, S., *et al.* 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**: 665-668.
- PASSALACQUA, K.D., VARADARAJAN, A., ONDOV, B.D., OKOU, D.T., ZWICK, M.E. & BERGMAN, N.H. 2009. Structure and complexity of a bacterial transcriptome. *Journal of Bacteriology* **191**: 3203-3211.
- PEARSON, W.R. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods in Molecular Biology* **132**: 185-219.
- PEDDIREDDI, L., CHENG, C. & GANTA, R.R. 2009. Promoter analysis of macrophage- and tick cell-specific differentially expressed *Ehrlichia chaffeensis* p28-Omp genes. *BMC Microbiology* **9**: 99.
- PERKINS, T.T., KINGSLEY, R.A., FOOKES, M.C., GARDNER, P.P., JAMES, K.D., YU, L., ASSEFA, S.A., HE, M., CROUCHER, N.J., PICKARD, D.J., *et al.* 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella* Typhi. *PLoS Genetics* **5**: e1000569.
- PETES, T.D. & HILL, C.W. 1988. Recombination between repeated genes in microorganisms. *Annual Review of Genetics* **22**: 147-168.
- PETNEY, T.N., HORAK, I.G. & RECHAV, Y. 1987. The ecology of the African vectors of heartwater, with particular reference to *Amblyomma hebraeum* and *Amblyomma variegatum*. *Onderstepoort Journal of Veterinary Research* **54**: 381-395.
- PIZZA, M., SCARLATO, V., MASIGNANI, V., GIULIANI, M.M., ARICÒ, B., COMANDUCCI, M., JENNINGS, G.T., BALDI, L., BARTOLINI, E., CAPECCHI, B., *et al.* 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**: 1816-1820.

- POP, M. 2009. Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics* **10**: 354-366.
- POSTIGO, M., TAOUFIK, A., BELL-SAKYI, L., DE VRIES, E., MORRISON, W.I. & JONGEJAN, F. 2007. Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. *Veterinary Microbiology* **122**: 298-305.
- POWELL, K. 2004. DNA vaccines – back in the saddle again? *Nature Biotechnology* **22**: 799-801.
- PRESTON, P.M., BROWN, C.G.D. & RICHARDSON, W. 1992. Cytokines inhibit the development of trophozoite-infected cells of *Theileria annulata* and *Theileria parva* but enhance the proliferation of macrophage-infected cell lines. *Parasite Immunology* **14**: 125-141.
- PRETORIUS, A., COLLINS, N.E., STEYN, H.C. & ALLSOPP, B.A. 2002a. Sequence analysis of three *Ehrlichia ruminantium* LambdaGEM-11 clones. *Annals of the New York Academy of Sciences* **969**: 155-158.
- PRETORIUS, A., VAN STRIJP, F., BRAYTON, K.A., COLLINS, N.E. & ALLSOPP, B.A. 2002b. Genetic immunization with *Ehrlichia ruminantium* GroEL and GroES homologues. *Annals of the New York Academy of Sciences* **969**: 151-154.
- PRETORIUS, A., COLLINS, N.E., STEYN, H.C., VAN STRIJP, F., VAN KLEEF, M. & ALLSOPP, B.A. 2007. Protection against heartwater by DNA immunisation with four *Ehrlichia ruminantium* open reading frames. *Vaccine* **25**: 2316-2324.
- PRETORIUS, A., VAN KLEEF, M., COLLINS, N.E., TSHIKUDO, N., LOUW, E., FABER, F.E., VAN STRIJP, M.F. & ALLSOPP B.A. 2008. A heterologous prime/boost immunisation strategy protects against virulent *E. ruminantium* Welgevonden needle challenge but not against tick challenge. *Vaccine* **26**: 4363-4371.
- PRETORIUS, A., LIEBENBERG, J., LOUW, E., COLLINS, N.E. & ALLSOPP B.A. 2010. Studies of a polymorphic *E. ruminantium* gene for use as a component of a recombinant vaccine against heartwater. *Vaccine* **28**: 3531-3539.
- PROVOST, A. & BEZUIDENHOUT, J.D. 1987. The historical background and global importance of heartwater. *Onderstepoort Journal of Veterinary Research* **54**: 165-169.
- PROZESKY, L. & DU PLESSIS, J.L. 1987. Heartwater. The development and life cycle of *Cowdria ruminantium* in the vertebrate host, ticks and cultured endothelial cells. *Onderstepoort Journal of Veterinary Research* **54**: 193-196.

- QI, W., KÄSER, M., RÖLTGEN, K., YEBOAH-MANU, D. & PLUSCHKE, G. 2009. Genomic diversity and evolution of *Mycobacterium ulcerans* revealed by next-generation sequencing. *PLoS Pathogens* **5**: e1000580.
- RAFATI, S., ZAHEDIFARD, F. & NAZGOUEE, F. 2006. Prime-boost vaccination using cysteine proteinases type I and II of *Leishmania infantum* confers protective immunity in murine visceral leishmaniasis. *Vaccine* **24**: 2169-2175.
- RAMARAO, N., GRAY-OWEN, S.D., BACKERT, S. & MEYER, T.F. 2000. *Helicobacter pylori* inhibits phagocytosis by professional phagocytes involving type IV secretion components. *Molecular Microbiology* **37**: 1389-1404.
- RAPP, U.K. & KAUFMANN, S.H.E. 2004. DNA vaccination with gp96-peptide fusion proteins induces protection against an intracellular bacterial pathogen. *International Immunology* **16**: 597-605.
- RAPPUOLI, R. 2000. Reverse vaccinology. *Current Opinion in Microbiology* **3**: 445-450.
- RAPPUOLI, R. 2001. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* **19**: 2688-2691.
- RAPPUOLI, R. 2007. Bridging the knowledge gaps in vaccine design. *Nature Biotechnology* **25**: 1361-1366.
- REDDY, G.R. 1995. Determining the sequence of parasite DNA. *Parasitology Today* **11**: 37-42.
- RILEY, M. 1993. Functions of the gene products of *Escherichia coli*. *Microbiological Reviews* **57**: 862-952.
- ROCHA, E.P. 2003. An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: from duplications to genome reduction. *Genome Research* **13**: 1123-1132.
- ROITT, I. 1991. Essential immunology. 8th ed. Oxford: Blackwell Scientific Publications, pp 3-20.
- RONAGHI, M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Research* **11**: 3-11.
- ROSET, M.S., CIOCCHINI, A.E., UGALDE, R.A. & IÑÓN DE IANNINO, N. 2004. Molecular cloning and characterization of *cgt*, the *Brucella abortus* cyclic beta-1,2-glucan transporter gene, and its role in virulence. *Infection and Immunity* **72**: 2263-2271.

- ROSS, B.C., CZAJKOWSKI, L., HOCKING, D., MARGETTS, M., WEBB, E., ROTHEL, L., PATTERSON, M., AGIUS, C., CAMUGLIA, S., REYNOLDS, E., *et al.* 2001. Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. *Vaccine* **19**: 4135-4142.
- ROSSOUW, M., NEITZ, A.W., DE WAAL, D.T., DU PLESSIS, J.L., VAN GAS, L. & BRETT, S. 1990. Identification of the antigenic proteins of *Cowdria ruminantium*. *Onderstepoort Journal of Veterinary Research* **57**: 215-221.
- RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.-A. & BARRELL, B. 2000. Artemis: sequence visualisation and annotation. *Bioinformatics* **16**: 944-945.
- SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.
- SAMBROOK, J., FRITSCH, E.H. & MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp E3-E4.
- SANGER, F., NICKLEN, S. & COULSON, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.
- SEDEGAH, M., CHAROENVIT, Y., MINH, L., BELMONTE, M., MAJAM, V.F., ABOT, S., GANESHAN, H., KUMAR, S., BACON, D.J., STOWERS, A., *et al.* 2004. Reduced immunogenicity of DNA vaccine plasmids in mixtures. *Gene Therapy* **11**: 448-456.
- SEGAL, E.D., CHA, J., LO, J., FALKOW, S. & TOMPKINS, L.S. 1999. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 14559-14564.
- SERRUTO, D. & RAPPUOLI, R. 2006. Post-genomic vaccine development. *FEBS Letters* **580**: 2985-2992.
- SERRUTO, D., SERINO, L., MASIGNANI, V. & PIZZA, M. 2009. Genome-based approaches to develop vaccines against bacterial pathogens. *Vaccine* **27**: 3245-3250.
- SHE, Q., CONFALONIERI, F., ZIVANOVIC, Y., MEDINA, N., BILLAULT, A., AWAYEZ, M.J., THINGOC, H.P., PHAM, B.T., VAN DER OOST, J., DUGUET, M. & GARRETT, R.A. 2000. A BAC library and paired-PCR approach to mapping and completing the genome sequence of *Sulfolobus solfataricus* P2. *DNA Sequence* **11**: 183-192.

- SHENDURE, J., PORRECA, G.J., REPPAS, N.B., LIN, X., MCCUTCHEON, J.P., ROSENBAUM, A.M., WANG, M.D., ZHANG, K., MITRA, R.D. & CHURCH, G.M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728-1732.
- SIGRIST, C.J.A., CERUTTI, L., HULO, N., GATTIKER, A., FALQUET, L., PAGNI, M., BAIROCH, A. & BUCHER, P. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in Bioinformatics* **3**: 265-274.
- SIMPSON, B.C., LINDSAY, M.S., MORRIS, J.R., MUIRHEAD, F.S., POLLOCK, A., PRICHARD, S.G., STANLEY, H.G., THIRLWELL, G.R., HUNTER, A.G., BRADLEY, J., *et al.* 1987. Protection of cattle against heartwater in Botswana: comparative efficacy of different methods against natural and blood-derived challenges. *The Veterinary Record* **120**: 135-138.
- SINGU, V., LIU, H., CHENG, C. & GANTA, R.R. 2005. *Ehrlichia chaffeensis* expresses macrophage- and tick cell-specific 28-kilodalton outer membrane proteins. *Infection and Immunity* **73**: 79-87.
- SINGU, V., PEDDIREDDI, L., SIRIGIREDDY, K.R., CHENG, C., MUNDERLOH, U. & GANTA, R.R. 2006. Unique macrophage and tick cell-specific protein expression from the p28/p30-outer membrane protein multigene locus in *Ehrlichia chaffeensis* and *Ehrlichia canis*. *Cellular Microbiology* **8**: 1475-1487.
- SONNHAMMER, E.L. & DURBIN, R. 1994. A workbench for large-scale sequence homology analysis. *Computer Applications in the Biosciences* **10**: 301-307.
- SONNHAMMER, E.L. & DURBIN, R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1-10.
- SPRENGER, J., FINK, J.L. & TEASDALE, R.D. 2006. Evaluation and comparison of mammalian subcellular localization prediction methods. *BMC Bioinformatics* **7**: S3.
- SPREULL, J. 1904. Heartwater inoculation experiments. *Agricultural Journal of the Cape of Good Hope* **24**: 433-442.
- STADEN, R., BEAL, K.F. & BONFIELD, J.K. 2000. The Staden package, 1998. *Methods in Molecular Biology* **132**: 115-130.
- STEMKE-HALE, K., KALTENBOECK, B., DEGRAVES, F.J., SYKES, K.F., HUANG, J., BU, C.H. & JOHNSTON, S.A. 2005. Screening the whole genome of a pathogen *in vivo* for individual protective antigens. *Vaccine* **23**: 3016-3025.

- STEPHENS, R.S., KALMAN, S., LAMMEL, C., FAN, J., MARATHE, R., ARAVIND, L., MITCHELL, W., OLINGER, L., TATUSOV, R.L., ZHAO, Q., *et al.* 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754-759.
- STEYN, H.C., PRETORIUS, A., MCCRINDLE, C.M.E., STEINMANN, C.M.L. & VAN KLEEF, M. 2008. A quantitative real-time PCR assay for *Ehrlichia ruminantium* using pCS20. *Veterinary Microbiology* **131**: 258-265.
- STORM, A.J., STORM, C., CHEN, J., ZANDBERGEN, H., JOANNY, J.F. & DEKKER, C. 2005. Fast DNA translocation through a solid-state nanopore. *Nano Letters* **5**: 1193-1197.
- STOTHARD, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**: 1102-1104.
- STRAUSS, E.C., KOBORI, J.A., SIU, G. & HOOD, L.E. 1986. Specific-primer-directed DNA sequencing. *Analytical Biochemistry* **154**: 353-360.
- SUNDQUIST, A., RONAGHI, M., TANG, H., PEVZNER, P. & BATZOGLOU, S. 2007. Whole-genome sequencing and assembly with high-throughput, short-read technologies. *PLoS ONE* **2**: e484.
- SULSONA, C.R., MAHAN, S.M. & BARBET, A.F. 1999. The *map1* gene of *Cowdria ruminantium* is a member of a multigene family containing both conserved and variable genes. *Biochemical and Biophysical Research Communications* **257**: 300-305.
- SYKES, K.F. & JOHNSTON, S.A. 1999. Genetic live vaccines mimic the antigenicity but not pathogenicity of live viruses. *DNA and Cell Biology* **18**: 521-531.
- TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B.B., SIDDIQUI, A., *et al.* 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**: 377-382.
- TETTELIN, H., SAUNDERS, N.J., HEIDELBERG, J., JEFFRIES, A.C., NELSON, K.E., EISEN, J.A., KETCHUM, K.A., HOOD, D.W., PEDEN, J.F., DODSON, R.J., *et al.* 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**: 1809-1815.
- The Gene Ontology Consortium. 2000. Gene Ontology: Tool for the unification of biology. *Nature Genetics* **25**: 25-29.

- THOMPSON, D.V., MELCHERS, L.S., IDLER, K.B., SCHILPEROORT, R.A. & HOOYKAAS, P.J. 1988. Analysis of the complete nucleotide sequence of the *Agrobacterium tumefaciens* virB operon. *Nucleic Acids Research* **16**: 4621-4636.
- THOMPSON, J.D., GIBSON, T.J. & HIGGINS, D.G. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* **2**: Unit 2.3.
- TOTTÉ, P., BLANKAERT, D., ZILIMWABAGABO, P. & WERENNE, J. 1993. Inhibition of *Cowdria ruminantium* infectious yield by interferons alpha and gamma in endothelial cells. *Revue d'Élevage et de Médecine Vétérinaire des Pays Tropicaux* **46**: 189-194.
- TOTTÉ, P., VACHIERY, N., MARTINEZ, D., TRAP, I., BALLINGALL, K.T., MACHUGH, N.D., BENSAID, A. & WERENNE, J. 1996. Recombinant bovine interferon gamma inhibits the growth of *Cowdria ruminantium* but fails to induce major histocompatibility complex class II following infection of endothelial cells. *Veterinary Immunology and Immunopathology* **53**: 61-71.
- TOTTÉ, P., MCKEEVER, D., MARTINEZ, D. & BENSAID, A. 1997. Analysis of T-cell responses in cattle immunized against heartwater by vaccination with killed elementary bodies of *Cowdria ruminantium*. *Infection and Immunity* **65**: 236-241.
- TOTTÉ, P., BENSAID, A., MAHAN, S.M., MARTINEZ, D. & MCKEEVER, D.J. 1999. Immune responses to *Cowdria ruminantium* infections. *Parasitology Today* **15**: 286-290.
- TURNER, D.J., KEANE, T.M., SUDBERY, I. & ADAMS, D.J. 2009. Next-generation sequencing of vertebrate experimental organisms. *Mammalian Genome* **20**: 327-338.
- UILENBERG, G. 1982. Experimental transmission of *Cowdria ruminantium* by the Gulf coast tick *Amblyomma maculatum*: danger of introducing heartwater and benign African Theileriasis onto the American mainland. *American Journal of Veterinary Research* **43**: 1279-1282.
- UILENBERG, G. 1983. Heartwater (*Cowdria ruminantium* infection): current status. *Advances in Veterinary Science and Comparative Medicine* **27**: 427-480.
- UILENBERG, G. 1990. Extension de la tique *Amblyomma variegatum* dans les Antilles: comment expliquer cette grave menace et que faire? *Revue d'Élevage et de Médecine Vétérinaire des Pays Tropicaux* **43**: 297-299.
- ULMER, J.B., WAHREN, B. & LIU, M.A. 2006. Gene-based vaccines: recent technical and clinical advances. *Trends in Molecular Medicine* **12**: 216-222.

- VACHIÉRY, N., LEFRANÇOIS, T., ESTEVES, I., MOLIA, S., SHEIKBOUDOU, C., KANDASSAMY, Y. & MARTINEZ, D. 2006. Optimisation of the inactivated vaccine dose against heartwater and in vitro quantification of *Ehrlichia ruminantium* challenge material. *Vaccine* **24**: 4747-4756.
- VAN BELKUM, A., SCHERER, S., VAN ALPHEN, L. & VERBRUGH, H. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiology and Molecular Biology Reviews* **62**: 275-293.
- VANBUSKIRK, K.M., O'NEILL, M.T., DE LA VEGA, P., MAIER, A.G., KRZYCH, U., WILLIAMS, J., DOWLER, M.G., SACCI, J.B. Jr., KANGWANRANGSAN, N., TSUBOI, T., *et al.* 2009. Preerythrocytic, live-attenuated *Plasmodium falciparum* vaccine candidates by design. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 13004-13009.
- VAN DE PYPEKAMP, H.E. & PROZESKY, L. 1987. Heartwater. An overview of the clinical signs, susceptibility and differential diagnoses of the disease in domestic ruminants. *Onderstepoort Journal of Veterinary Research* **54**: 263-266.
- VAN HEERDEN, H., COLLINS, N.E., ALLSOPP, M.T.E.P. & ALLSOPP, B.A. 2002. Major outer membrane proteins of *Ehrlichia ruminantium* encoded by multigene family. *Annals of the New York Academy of Sciences* **969**: 131-134.
- VAN HEERDEN, H., COLLINS, N.E., BRAYTON, K.A., RADEMEYER, C. & ALLSOPP, B.A. 2004a. Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium*. *Gene* **330**: 159-168.
- VAN HEERDEN, H., STEYN, H.C., ALLSOPP, M.T.E.P., ZWEYGARTH, E., JOSEMANS, A.I. & ALLSOPP, B.A. 2004b. Characterization of the pCS20 region of different *Ehrlichia ruminantium* isolates. *Veterinary Microbiology* **101**: 279-291.
- VAN KLEEF, M., GUNTER, N.J., MACMILLAN, H., ALLSOPP, B.A., SHKAP, V. & BROWN, W.C. 2000. Identification of *Cowdria ruminantium* antigens that stimulate proliferation of lymphocytes from cattle immunized by infection and treatment or with inactivated organisms. *Infection and Immunity* **68**: 603-614.
- VAN KLEEF, M., MACMILLAN, H., GUNTER, N.J., ZWEYGARTH, E., ALLSOPP, B.A., SHKAP, V., DU PLESSIS, D.H. & BROWN, W.C. 2002. Low molecular weight proteins of *Cowdria ruminantium* (Welgevonden isolate) induce bovine CD4⁺-enriched T-cells to proliferate and produce interferon- γ . *Veterinary Microbiology* **85**: 259-273.

- VAN VLIET, A.H., JONGEJAN, F., VAN KLEEF, M. & VAN DER ZEIJST, B.A. 1994. Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infection and Immunity* **62**: 1451-1456.
- VAN VLIET, A.H. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters* **302**: 1-7.
- VENTER, J.C., ADAMS, M.D., MYERS, E.W., LI, P.W., MURAL, R.J., SUTTON, G.G., SMITH, H.O., YANDELL, M., EVANS, C.A., HOLT, R.A., *et al.* 2001. The sequence of the human genome. *Science* **291**: 1304-1351.
- VOELKERDING, K.V., DAMES, S.A. & DURTSCHI, J.D. 2009. Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry* **55**: 641-658.
- VON HEIJNE, G. 1985. Signal sequences. The limits of variation. *Journal of Molecular Biology* **184**: 99-105.
- VON HEIJNE, G. 1992. Membrane protein structure prediction – Hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology* **225**: 487-494.
- VOYICH, J.M., STURDEVANT, D.E., BRAUGHTON, K.R., KOBAYASHI, S.D., LEI, B., VIRTANEVA, K., DORWARD, D.W., MUSSER, J.M. & DELEO, F.R. 2003. Genome-wide protective response used by group A *Streptococcus* to evade destruction by human polymorphonuclear leukocytes. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 1996-2001.
- WAGHELA, S.D., RURANGIRWA, F.R., MAHAN, S.M., YUNKER, C.E., CRAWFORD, T.B., BARBET, A.F., BURRIDGE, M.J. & MCGUIRE, T.C. 1991. A cloned DNA probe identifies *Cowdria ruminantium* in *Amblyomma variegatum* ticks. *Journal of Clinical Microbiology* **29**: 2571-2577.
- WAKEEL, A., KURIAKOSE, J.A. & MCBRIDE, J.W. 2009. An *Ehrlichia chaffeensis* tandem repeat protein interacts with multiple host targets involved in cell signaling, transcriptional regulation, and vesicle trafficking. *Infection and Immunity* **77**: 1734-1745.
- WALKER, J.B. & OLWAGE, A. 1987. The tick vectors of *Cowdria ruminantium* (Ixodoidea, Ixodidae, genus *Amblyomma*) and their distribution. *Onderstepoort Journal of Veterinary Research* **54**: 353-379.

- WANG, Q.M., SUN, S.H., HU, Z.L., YIN, M., XIAO, C.J. & ZHANG, J.C. 2004a. Improved immunogenicity of a tuberculosis DNA vaccine encoding ESAT6 by DNA priming and protein boosting. *Vaccine* **22**: 3622-3627.
- WANG, R., EPSTEIN, J., CHAROENVIT, Y., BARACEROS, F.M., RAHARDJO, N., GAY, T., BANANIA, J.G., CHATTOPADHYAY, R., DE LA VEGA, P., RICHIE, T.L., *et al.* 2004b. Induction in humans of CD8+ and CD4+ T cell and antibody responses by sequential immunization with malaria DNA and recombinant protein. *The Journal of Immunology* **172**: 5561-5569.
- WANG, J., WANG, W., LI, R., LI, Y., TIAN, G., GOODMAN, L., FAN, W., ZHANG, J., LI, J., ZHANG, J., *et al.* 2008. The diploid genome sequence of an Asian individual. *Nature* **456**: 60-65.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57-63.
- WANG, T. & YANG, J. 2009. Using the nonlinear dimensionality reduction method for the prediction of subcellular localization of Gram-negative bacterial proteins. *Molecular Diversity* **13**: 475-481.
- WARREN, R.L., SUTTON, G.G., JONES, S.J. & HOLT, R.A. 2006. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* **23**: 500-501.
- WEAVER, R.F. & HEDRICK, P.W. 1992. *Genetics*. 2nd ed. Dubuque: Wm. C. Brown Publishers, pp 296-302.
- WEISS, K.E., HAIG, D.A. & ALEXANDER, R.A. 1952. Aureomycin in the treatment of heartwater. *Onderstepoort Journal of Veterinary Research* **25**: 41-50.
- WEISS, A.A., JOHNSON, F.D. & BURNS, D.L. 1993. Molecular characterization of an operon required for pertussis toxin secretion. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 2970-2974.
- WEITZMANN, M.N., WOODFORD, K.J. & USDIN, K. 1997. DNA secondary structures and the evolution of hypervariable tandem arrays. *Journal of Biological Chemistry* **272**: 9517-9523.
- WHEELER, D.A., SRINIVASAN, M., EGHOLM, M., SHEN, Y., CHEN, L., MCGUIRE, A., HE, W., CHEN, Y.J., MAKHIJANI, V., ROTH, G.T., *et al.* 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**: 872-876.

- WHITEFORD, N., HASLAM, N., WEBER, G., PRÜGEL-BENNETT, A., ESSEX, J.W., ROACH, P.L., BRADLEY, M. & NEYLON, C. 2005. An analysis of the feasibility of short read sequencing. *Nucleic Acids Research* **33**: e171.
- WICKER, T., SCHLAGENHAUF, E., GRANER, A., CLOSE, T.J., KELLER, B. & STEIN, N. 2006. 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* **7**: 275.
- WILLIAMS, K.P., SOBRAL, B.W. & DICKERMAN, A.W. 2007. A robust species tree for the Alphaproteobacteria. *Journal of Bacteriology* **189**: 4578-4586.
- WILSON, C.G., KAJANDER, T. & REGAN, L. 2005. The crystal structure of NlpI. A prokaryotic tetratricopeptide repeat protein with a globular fold. *FEBS Journal* **272**: 166-179.
- WIUF, C. & HEIN, J. 2000. The coalescent with gene conversion. *Genetics* **155**: 451-462.
- WIZEMANN, T.M., HEINRICHS, J.H., ADAMOU, J.E., ERWIN, A.L., KUNSCH, C., CHOI, G.H., BARASH, S.C., ROSEN, C.A., MASURE, H.R., TUOMANEN, E., *et al.* 2001. Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infection and Immunity* **69**: 1593-1598.
- WOOSTER, R., BIGNELL, G., LANCASTER, J., SWIFT, S., SEAL, S., MANGION, J., COLLINS, N., GREGORY, S., GUMBS, C. & MICKLEM, G. 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**: 789-792.
- WU, M., SUN, L.V., VAMATHEVAN, J., RIEGLER, M., DEBOY, R., BROWNLIE, J.C., MCGRAW, E.A., MARTIN, W., ESSER, C., AHMADINEJAD, N., *et al.* 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology* **2**: 0327-0341.
- XIA, Q., GUO, Y., ZHANG, Z., LI, D., XUAN, Z., LI, Z., DAI, F., LI, Y., CHENG, D., LI, R., *et al.* 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* **326**: 433-436.
- YANG, X., HUDSON, M., WALTERS, N., BARGATZE, R.F. & PASCUAL, D.W. 2005. Selection of protective epitopes for *Brucella melitensis* by DNA vaccination. *Infection and Immunity* **73**: 7297-7303.
- YU, C.S., LIN, C.J. & HWANG, J.K. 2004. Predicting subcellular localisation of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science* **13**: 1402-1406.

- YU, N.Y., WAGNER, J.R., LAIRD, M.R., MELLI, G., REY, S., LO, R., DAO, P., SAHINALP, S.C., ESTER, M., FOSTER, L.J. & BRINKMAN, F.S. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**: 1608-1615.
- YUAN, Z., DAVIS, M.J., ZHANG, F. & TEASDALE, R.D. 2003. Computational differentiation of N-terminal signal peptides and transmembrane helices. *Biochemical and Biophysical Research Communications* **312**: 1278-1283.
- ZARLENGA, D.S. & GASBARRE, L.C. 2009. From parasite genomes to one healthy world: Are we having fun yet? *Veterinary Parasitology* **163**: 235-249.
- ZHANG, X., LUO, T., KEYSARY, A., BANETH, G., MIYASHIRO, S., STRENGER, C., WANER, T. & MCBRIDE, J.W. 2008a. Genetic and antigenic diversities of major immunoreactive proteins in globally distributed *Ehrlichia canis* strains. *Clinical and Vaccine Immunology* **15**: 1080-1088.
- ZHANG, C., XIONG, Q., KIKUCHI, T. & RIKIHISA, Y. 2008b. Identification of 19 polymorphic major outer membrane protein genes and their immunogenic peptides in *Ehrlichia ewingii* for use in a serodiagnostic assay. *Clinical and Vaccine Immunology* **15**: 402-411.
- ZHU, B., NETHERY, K.A., KURIAKOSE, J.A., WAKEEL, A., ZHANG, X. & MCBRIDE, J.W. 2009. Nuclear translocated *Ehrlichia chaffeensis* ankyrin protein interacts with a specific adenine-rich motif of host promoter and intronic Alu elements. *Infection and Immunity* **77**: 4243-4255.
- ZIENTZ, E., DANDEKAR, T. & GROSS, R. 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiology and Molecular Biology Reviews* **68**: 745-770.
- ZIVKOVIC, Z., ESTEVES, E., ALMAZÁN, C., DAFFRE, S., NIJHOF, A.M., KOCAN, K.M., JONGEJAN, F. & DE LA FUENTE, J. 2010. Differential expression of genes in salivary glands of male *Rhipicephalus (Boophilus) microplus* in response to infection with *Anaplasma marginale*. *BMC Genomics* **11**: 186.
- ZWEYGARTH, E. & JOSEMANS, A.I. 2001a. A chemically defined medium for the growth of *Cowdria ruminantium*. *Onderstepoort Journal of Veterinary Research* **68**: 37-40.
- ZWEYGARTH, E. & JOSEMANS, A.I. 2001b. Continuous *in vitro* propagation of *Cowdria ruminantium* (Welgevonden stock) in a canine macrophage-monocyte cell line. *Onderstepoort Journal of Veterinary Research* **68**: 155-157.

ZWEYGARTH, E., JOSEMANS, A.I., VAN STRIJP, M.F., LOPEZ-REBOLLAR, L., VAN KLEEF, M. & ALLSOPP, B.A. 2005. An attenuated *Ehrlichia ruminantium* (Welgevonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* **23**: 1695-1702.

ZWEYGARTH, E., JOSEMANS, A.I. & STEYN, H.C. 2008. Experimental use of the attenuated *Ehrlichia ruminantium* (Welgevonden) vaccine in Merino sheep and Angora goats. *Vaccine* **26S**: G34-39.

Appendix B: Materials, buffers, media and solutions

B1: Suppliers of the materials used in this study

Supplier	Product
Amersham Biosciences	pMOS <i>Blue</i> blunt ended cloning kit, TempliPhi DNA Sequencing Template Amplification Kit, [³ H] thymidine
Applied Biosystems	Amplitaq Gold polymerase, Dye-terminator Cycle Sequencing kit
BDH	Glycine, glycerol, magnesium chloride, potassium bi-phosphate
Bio-Rad	Biolistic [®] 1.6 Micron Gold, Criterion [™] XT precast gels, RC/DC Protein Assay, XT MOPS Running Buffer, XT Sample Buffer, XT Reducing Agent, Bio-Safe Coomassie
Celtic Molecular Diagnostics	mouse anti-bovine IFN- γ mAb CC302
Centaur	Eutha-Nase
Costar	Half-area 96-well plates
Gibco BRL Products	Concert Rapid PCR Purification kit
Highveldt biologicals	Foetal calf serum
Immonodiagnostik	rabbit anti-bovine IFN- γ anti-serum
Inqaba Biotec	Primers
Invitex	Invisorb [®] Spin Plasmid Mini <i>Two</i> kit, MSB [®] Spin PCRapace kit
Invitrogen	pET102/D-TOPO [®] expression system, Platinum [®] <i>px</i> DNA polymerase kit, RPMI-1640 + GlutaMAX-I
Macherey-Nagel	NucleoBond [®] Xtra Maxi purification Kit, Protino [®] Ni 1000 prepacked columns kit
Merck Biosciences	Acetone, calcium chloride, chloroform, citric acid, ethanol, glucose, isoamyl alcohol, methanol, potassium chloride, sodium carbonate, sodium mono-phosphate, tri-sodium citrate, tryptone, Tween-20
Millipore Corporation	dH ₂ O, PVDF membranes, MAIPS 4510 Multiscreen [™] -IP filtration plates
Montanide	ISA50
MP Biomedicals	Penicillin, streptomycin
Novagen	Overnight Express Autoinduction system 1, BugBuster [®] Protein Extraction Reagent
Ondestepoort Biological Products	Elsevers medium, PBS
Packard BioScience	Ultima gold F scintillation solution
Pfizer	Liquamycin/LA
Pierce	SuperSignal [®] West Pico Chemiluminescent substrate
Promega	Ethidium bromide solution, isopropyl- β -D-thiogalactopyranoside (IPTG), <i>Pfu</i> polymerase, pGEM-T Easy cloning kit, shrimp alkaline phosphatase, T4 ligase
Qiagen	QIAquick PCR Purification Kit, QIAprep 8 Miniprep Kit,
Roche	Anti-His ₆ antibodies, DNase I, High Pure PCR product purification kit, High Pure Plasmid Isolation kit, Restriction endonucleases (<i>Bam</i> HI, <i>Eco</i> RI, <i>Sal</i> I and <i>Xba</i> I)
Sigma	Ampicillin, anti-rabbit IgG alkaline phosphatase conjugate, bromophenol blue, concanavalin A, Dulbecco's PBS, Fast BCIP/NBT substrate tablets, Hank's Balanced Salt Solution, Histopaque [®] -1077, 2-mercaptoethanol, percoll, phenol, proteinase K, RNase, spermidine, trypan blue

Supplier	Product
Stratagene	Klenow Fill-In kit
TaKaRa Bio Inc.	TaKaRa Ex Taq, TaKaRa recochips
USB	Agar, boric acid, EDTA, HCl, magnesium sulphate, potassium bi-phosphate, SDS, sodium chloride, Tris base
Walac	96 well glass fibre filters
White Sci	Agarose
Zymed	HRP-goat-anti-mouse IgG

B2: Preparation of buffers, media and solutions

B2.1. Buffers for general laboratory use

Ampicillin

Prepare a stock solution of 10 µg/ml by dissolving 50 mg ampicillin in 5 ml dH₂O. The solution was filter sterilised and stored in aliquots at -20°C.

1M IPTG

Dissolve 1.19 g isopropyl-β-D-thiogalactopyranoside in 50 ml dH₂O. Sterilise by filtration, aliquot and store at -20°C.

LB Agar plates

Dissolve 10 g tryptone, 5 g yeast, 10 g NaCl and 15 g bacto-agar in 800 ml dH₂O. Adjust volume to 1000 ml with dH₂O and sterilise by autoclaving. Allow to cool to 55°C and add the appropriate amount of antibiotics before pouring the plates. Store at 4°C.

LB broth

Dissolve 10 g tryptone, 5 g yeast and 10 g NaCl in 800 ml dH₂O. Adjust volume to 1000 ml with dH₂O and sterilise by autoclaving. Store at 4°C and add appropriate antibiotics prior to use.

SOC

Buffer consists of 20 g tryptone, 5 g yeast extract, 0.5 g NaCl and 2.5 ml 1 M KCl. Adjust pH to 7 with NaOH, make up to 970 ml and autoclave. Add sterile 1 M MgCl₂ and 20 ml sterile glucose prior to use.

10x TBE

Dissolve 108 g Tris base, 55 g boric acid and 7.44 g EDTA in 800 ml dH₂O. Adjust volume to 1000 ml with dH₂O.

X-gal

Dissolve 400 mg 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) in 20 ml N,N'-dimethyl formamide. Aliquot and store at -20°C.

B2.2. Western blots buffers

Blocking buffer

Dissolve 1 g bovine serum albumin in 100 ml 1x PBS.

Transfer buffer

Dissolve 9.1 g Tris base (38 mM), 43.25 g glycine (288 mM) in 1700 ml dH₂O and add 300 ml methanol (pH ~8.3-8.4).

Wash buffer

Dissolve 500 µl Tween-20 in 1000 ml 1x PBS.

B2.3. ELISpot buffers

Blocking medium

RPMI-1640 medium supplemented with 10% heat inactivated foetal calf serum.

Coating antibody

Dissolve mouse anti-bovine IFN- γ mAb CC302 in sterile PBS (100 µg/ml), aliquot and store at -20°C.

Carbonate/bicarbonate coating buffer

Buffer consists of 15mM Na₂CO₃ and 35 mM NaHCO₃. Adjust pH to 9.6, and filter sterilize.

Complete RPMI-1640 medium

Add to RPMI-1640 + GlutaMAX-1: 10% foetal calf serum, 5 x 10⁻⁵ M mercaptoethanol, 100 U/ml penicillin and 0.1 mg/ml streptomycin.

dH₂O-T

dH₂O with 0.05% Tween-20.

PBS-T

PBS with 0.05% Tween-20.

PBS-T/BSA

PBS-T with 0.1% bovine serum albumin

Rabbit anti-bovine IFN- γ antiserum

Dissolve rabbit anti bovine IFN- γ antibody in 100 µl dH₂O and sterilise through a 0.22 µm filter.

Monoclonal anti-rabbit IgG alkaline phosphate conjugate

Dissolve 1 tablet/10 ml dH₂O at room temperature for 30 min and sterilise through a 0.22 µm filter.

Appendix C: Primers

C1: Primers used to complete the genome sequence.

Primer name	Position	Primer sequence (5' > 3')
WTHIN992_1F	107 > 127	GGATTAGGACGTATAGGAAGG
WTHIN992_1R	561 > 539	CAGTTTGTGTACATGATCCAGC
WTHIN2864_27R	814 > 793	GGAAACATTTGGTGTGGTACG
WGAP49_2F	1050 > 1072	TGAGAATGCTGGATATAGTGTGC
WGAP49_F	1315 > 1350	TTTTTACAATGATTTATATAACCCTATAACAACAG
WGAP49walk_1F	1556 > 1583	ACAGAAAAGTAACACCTAATAACAATGC
WGAP49walk_1R	2467 > 2436	TGTTTGTGGGTATTGTAGTTAATTATATAGC
WGAP49_2R	2858 > 2835	TTGTCCAACCATAAAAATCTGTACG
1740_R	3321 > 3298	TTAGAGAAGCTGTGTTAGTTATGC
WGAP49_R	3331 > 3308	GGTTCAGAAATTAGAGAAGCTGTG
WTHIN24731_1F	5395 > 5420	CCAGATGTAATAGAAAATAGATGCAGC
WTHIN24731_1R	6977 > 6956	CTCTCCCTACATTTCTGTAGC
WTHIN24731_2F	8171 > 8197	TCACATGAAACATAACATTCAATTCCC
WTHIN24731_2R	9741 > 9717	AGATCTTGCTTTGCAAATAAATCGC
WTHIN24731_3F	12680 > 12707	GGCATTATTTACAACATGTATTAATGGC
WTHIN24731_3R	13900 > 13874	CTTGTTAATGATGGTAGATCTATACGC
WTHIN24731_4R	15023 > 15003	CACCAATGCTTGGATTTTCCC
WTHIN24731_5F	22812 > 22837	GCATCTAAGTTGAAACGTTTATCGG
WTHIN24731_5R	23272 > 23246	AGAAGATAAAAAGGATTGCTGTAATCG
WTHIN_29208_F	29208 > 29228	TGGTACAATGTTCAAGGAGGC
WGAP50_F	30429 > 30449	CTGTTTGCATCATCATGGTGG
1740_F	30466 > 30485	CACCACACTCATCATTGTCC
WTHIN_30489_R	30489 > 30468	ACAAGGACAATGATGAGTGTGG
WGAP50_R	30807 > 30780	GCACTGAAAATCAAATATAGACAGAACC
WTHIN24731_6F	39335 > 39356	TGTAACCTTTGGGTGAGATTGC
WTHIN24731_6R	40585 > 40561	AGTCGAAGGAATACTATTTGTAGCC
WTHIN_40819_R	40819 > 40795	AACAGAGGTATTACAATACATTGGC
WGAP51_F	43727 > 43758	AATCAATTAGTGGATATTATAGATACTGATGG
WGAP51_R	44893 > 44869	CAACTGGTACATTCAAATCTACACG
WTHIN_45092_R	45092 > 45067	CCATCTACAAATAGCACATTAGTTGC
WTHIN_45283_R	45283 > 45261	CTATAGATGCATGTTTACGGTGC
WTHIN_45361_F	45361 > 45385	TCGAATAGTGATAAACCAGTAGGTG
WGAP52_2F	45391 > 45414	GTAGGAGGTGCAAAAATATCAACG
WGAP52_F	45865 > 45886	GGTGGAGGTGACAGTATATGTG
440_LE	47113 > 47083	CGAAGGATAAGTATTATAGTAAATTGACAGC
WGAP52_R	48262 > 48236	TTCCAAGTCATATAGTCATATTTCCAG
WGAP52_2R	48382 > 48361	AAGGTCTGTTTGTATCAGTCTCG
WGAP53_F	57820 > 57849	TTTGTAGTAATTCTGTTACCATCTAATGTG
WGAP53_1491R	59015 > 58987	CTATATCTCATGTCACAACATCATATGC
WTHIN_59186_R	59186 > 59160	CAACAACAATTTGCTATTCTTGTTCAC
WTHIN1491_1F	60579 > 60605	AGTATTACACATAGGAAGAGTAAGCTG
WTHIN1491_1R	60811 > 60788	TCAACAGTTGCAAAAATTTATCCCC
WGAP53_1491F	61996 > 62019	TGCTTGTGATTTTATTGTGACAGG
WTHIN440_1R	62001 > 61976	CAAGCACTTCTTGTTAATAGCTATCG
WGAP53_R	62593 > 62570	AGTCAACATCCTATCTCAAGACTG
WTHIN14674_1F	67360 > 67384	AGTAATTGCTAATCAATGCAAGAGG
WTHIN14674_1R	67682 > 67659	TCCAATACAGTTAACTTACAGCCC
WTHIN14674_2F	79317 > 79344	AATATATCTGGTCTTGTTCACATAAGC
WTHIN14674_2R	79903 > 79879	AAAGAACCAACCTTAACTAATGAC
WGAP59_F	80643 > 80670	ATATTAAGCAGCTAAAGTCTTTGTAGAG
WGAP59walk_1F	81166 > 81195	TGGTATGTTAAAGATAAGATGGTTATTGAG
WGAP59walk_1R	81770 > 81744	TTTTACGATAGACATCATAACCACATG
WGAP59_R	82364 > 82335	CATATAACACAATAGTAATAAGGAGGACAC
WTHIN440_2F	87528 > 87550	TCCCTCCAGATTTGATTAGATGC



Primer name	Position	Primer sequence (5' > 3')
WTHIN440_2R	88170 > 88145	AGAGACTTACATTTACATCCACATCC
WTHIN_88855_R	88855 > 88834	CCAGCATGATGAGTTAATACGC
WGAP60_F	92316 > 92344	ACTGATGAAATGATAGAGATACTAACTCG
WGAP60_R	93556 > 93528	GCATTCGGTTAATCATAGTTTATTACATGC
WTHIN440_3F	111299 > 111325	TCTATGAAAGATAAGCTAAGTGATGGG
13618_RI	111763 > 111789	GTGTAATATAAGTGTCAACTGAAGGAG
15457_LE	112323 > 112301	CAACAACAGATCTAAGCTGAACC
WTHIN15457_1F	112728 > 112756	GTATTGTAATACAGAAGCTCAAGTATCTG
WTHIN15457_1R	112977 > 112954	ACCTAATACAGAACTATCAGCACC
WTHIN440_4F	116709 > 116730	GTGCATCAAGTACATCAGAAGC
WTHIN440_4R	117416 > 117390	ACATCCATATAAGTCTCTTATCACAGC
15457_RI	123395 > 123417	TGCAAAATAGAAGGAGAAGTGGG
WTHIN440_5F	124111 > 124136	CTTGAAAATCAACTTGATGATGGG
15288_LE	124137 > 124114	ACCCATCATACAAGTTGATTTTCC
WTHIN440_5R	124906 > 124882	AGCTACCTTTGACATTTATACCTCG
WTHIN15288_1F	127329 > 127351	TGTAAGTGTGATACTTGGAGTG
WTHIN15288_1R	128211 > 128185	TCTACCTAATACAATAACAACAAGCG
WTHIN15288_2F	129703 > 129725	CTGCAGTTATGATAAGCAAGGTG
WTHIN15288_2R	130292 > 130270	GCATAGCAAACACTACAGTCACAGC
WTHIN440_6F	134161 > 134183	GAAAGGAAACAATGACATGGGAG
WTHIN_1F	134315 > 134343	GTGGAATATTTTAAATAATGGACAAGATGC
WTHIN_1R	135640 > 135617	AACATGTCTATATGTAGTTGCC
WGAP19_F	138617 > 138645	CAAATATTGTATTGATAATTCAGTGTCC
WGAP19walk_1R	139723 > 139702	CATCCATTAGTAACCATGCTGC
WGAP19walk_2R	140593 > 140567	GATTTTCAGGTAATATGAAGAATGACGG
WGAP19_R	142129 > 142105	CTGATGACATCAGGTCTTTATTGTC
WTHIN912_1F	178258 > 178282	CTACATTGCACACATACATCATAGG
WTHIN912_1R	178733 > 178708	AGATGATAGATTGAAGACCTTAGCAC
WTHIN_183502_F	183489 > 183511	GATGTTCTACAGTAACCAAAGC
WTHIN_184516_R	184503 > 184482	CACATGCATGAACTACAACCTGG
WTHIN912_2F	194799 > 194821	TGGTAAAGTGAACCTTCAAGTGC
WTHIN912_2R	195553 > 195533	GACAGGAAATAACAAGGCTGC
WGAP18_F	199220 > 199246	TATTGACATTCATTCGGAAATATGGG
WGAP18walk_1F	199598 > 199619	TGTTTGGTAAAGTGTAGGAGAGG
WGAP18walk_1R	200616 > 200590	TCCAATCATATCAAAACACAACATCAG
WGAP17_F	201214 > 201189	CCACCTAAATCTTCATCATTGATACC
WTHIN_204439_F	204426 > 204448	GGTGAACCACTTGTAACATTGC
WTHIN24663_23R	205295 > 205320	GTTAAAAGTAGGACTGCTGTATTTGG
WTHIN24663_23F	206059 > 206037	TGCACAAGTCTAACAAGTCACTC
WTHIN24663_22R	208735 > 208759	TGTTGATGTAGGATTTTGTATGGC
WTHIN24663_22F	209611 > 209585	ATACTCGTTAACACTTATTCTAAAGCC
WGAP18_R	217387 > 217409	CTTTGAGCTATTAATGGTACGGC
WGAP17_R	218500 > 218471	ACATTTCAAAGATAACAAATCACAATATCC
WTHIN24663_20R	225283 > 225311	GTTATGTTATATCTATGTGCGGTTTATGG
WTHIN24663_20F	226046 > 226019	ACTAGATTCACACAATACATATCTCTCC
WTHIN24663_19R	230805 > 230829	TGCTCATACTTTTGAAATTCAGTCC
WTHIN24663_19F	231296 > 231269	GTAATACTAGAAGAATTATGCACTGTCC
WGAP37_F	236390 > 236414	TCATGTAGGAAAGTTTTGTGTTGTG
WGAP37_R	237392 > 237361	ACCATAGAACATTTCTTTAGTAGTTATATCC
18484_RCF	241513 > 241539	GTAGTGATAGGTTTGTAGTGTTAAGTG
WTHIN24663_15R	243146 > 243172	GATTTGTAGTTTTGGTCATACATGAGC
WTHIN_243656_F	243643 > 243665	TGGTATGTGAGTATTGCGATTGG
WTHIN24663_15F	243666 > 243644	ACCAATCGCAATACTCACATACC
WTHIN_244515_R	244503 > 244483	GCAGCAAGCTATCAAGACAGG
18484_RCR	244644 > 244614	GATAGTACCATATTCTCTATCATACTTACTG
WTHIN24663_13R	255834 > 255860	CCCTGGTTTATCTAAATATGGTTTTGC
WTHIN24663_13F	256412 > 256386	AGGAGAAACAATGATTGTATTAATGGC
WGAP38_F	269703 > 269731	ATAGCTATGTTATAAGGTGTAATTGAGTG
WGAP38_R	270006 > 269979	CACATTACCTTTTGCAACTTATAAACAC
WTHIN24663_11R	270649 > 270678	GTTAGTTAGAAGTAGTCTGATAACAATTCC



Primer name	Position	Primer sequence (5' > 3')
WTHIN24663_11F	271698 > 271674	GCTGCATCAGTATATCTTTTCATCAC
WGAP39_F	274371 > 274393	TCGGATACACTAAGAACAACACTGC
WGAP39_R	274762 > 274742	CGCTATCTGGAACCTTAGCAGG
WTHIN_276069_F	276055 > 276077	ACACTATGCTCTCTATGTGATGC
WTHIN_276967_R	276953 > 276930	CAGAGTTGCTATATCCCTATCCGC
WGAP40_F	279398 > 279424	CACTGTAAGTTTTGGTATTTAGATGGG
WGAP40_R	279840 > 279811	ATTTGTAGCATATAATACTATCAGTAGCAG
24993_RCF	282977 > 282999	CGATCTATGTCTCAAGGTAGAGC
WGAP41_F	283515 > 283539	TGAAGAGATGCTATCGTTAGTTGAG
WGAP41_R	283925 > 283905	AATATCCCAGCATTATCCCCC
24993_RCR	284382 > 284361	AAACTATGGCAGGAGTGATAGG
WGAP42_F	285508 > 285528	TGCTGATACAGTAGATGCTGC
WGAP42walk_1R	285933 > 285959	TTTTATGTCTTCTGTCTCTTCTATTGC
WGAP42_R	286888 > 286865	CTCCATCTTATCTACTAGTTCCGC
WTHIN24663_9R	295986 > 296011	GAAAGTGTATGCTGATGTATTAAGCC
WTHIN24663_9F	296960 > 296931	CATCATATCTAGTAACTTTAGGTAGCTCTC
WGAP103_R	299058 > 299085	TTATAATTCTATGTGGCTAGTCTTTTGG
WGAP103_F	299445 > 299418	GCACTTAAACACAATTGAAACTTTTGG
WTHIN24663_8F	300242 > 300218	GAAACACTTCATATACAGTACCCAC
WTHIN24663_7R	303780 > 303802	ACCTCTACTAAGACTGAGAGCAG
WTHIN24663_7F	304262 > 304239	CATATTTTCGACCTATTTCTGCCAC
WGAP102_R	313871 > 313894	AGCGATTTGTAATGTGTGAAACC
WGAP102walk_1R	314264 > 314288	TGCTTGATCAAATGAGATTGATTCCG
WGAP102walk_1F	315065 > 315035	GTACTATAGTTGAGATAACGAACATTAAGTC
849_F	315245 > 315209	ACAATTGATACCTAAGTAGCTACAGTC
WTHIN_315739_R	315726 > 315704	AAACTACCTACTGAACTACCAGC
WTHIN23036_1R	318174 > 318200	AGGCATTATTATTATGTTCTTGTGGG
WTHIN23036_1F	319336 > 319309	TTGTAGCATGTATTATTAGATCATCAGC
849_R	339626 > 339654	CTTGAACATACATACCACATATACCTACC
WGAP101_F	340493 > 340467	GAAGTTGTTATTGATGAAGTCATAGGG
WTHIN24663_5R	346090 > 346113	TCAGAAATAAGAGGTCATCGTAGG
WTHIN24663_5F	346592 > 346570	ACTGCCTTTCTCATAAGACTAGC
WTHIN24663_4R	348908 > 348931	AGACAACAGTATCTTGAGCATACG
WGAP100_R	349303 > 349332	GTGATTGTTAAGGATATTATTCTATGTTGC
WGAP100_F	349809 > 349784	GCAAGCTTGATACTTGTTAATCTGTC
WGAP99_R	350692 > 350716	TTGTAGTACAGATGGAAGGTAAGAC
WGAP99_F	351186 > 351160	CGATAAAAAGTTGAAACAACGATATCCC
WGAP98_R	351426 > 351450	TGAGGAAAGAGTTAGTATGCTTAGG
WGAP98_F	351691 > 351665	AGTACAATACATCGCTATAATAGGGTC
WGAP97_R	352556 > 352582	TTGTGTGGTCTTGTATTAATAGTTACG
WGAP97_F	352862 > 352840	AGATTCTGTCCATCATCATGAGC
WTHIN1639_1R	354298 > 354323	TGCAATAAAATATAAGGCATATGGGG
WGAP96_R	355056 > 355082	TGTGAGATAATGACTTAACAATATGGC
WGAP96_F	356303 > 356279	CTAAGCAGTATATTGGATCTTGACAG
WTHIN24663_3R	356595 > 356618	CAATATTCCTTGTGGTCTTAGAGC
WTHIN24663_3F	357305 > 357282	ACTCTTTACCCAAAGTAGTAACGC
WGAP95_R	360602 > 360623	AAAGGTACAGTAACTGGAGAGC
WGAP95_F	361685 > 361662	GTTATTACTCAGTACCGCATCTGG
WGAP94_R	362643 > 362668	CCATCATGACCATATAAGTACACTCC
WGAP94_F	364025 > 364001	ACAAGCTTTACTGTTCCATTTTCAG
WGAP16_F	371152 > 371178	ACATGACAAGATCTACAAAATCAAGAC
WGAP16_R	372094 > 372066	AGGAACAATTTGCAAATTCATAAATTGAG
WTHIN24663_1R	375188 > 375215	CAACATGTGTAATTGTTACAGATTGGAG
WTHIN24663_1F	376135 > 376113	TCTTGATAACAGATTGCCTCCTC
WGAP15_F	384394 > 384421	TGAATTGAAGACTTGATATGTATTCCCTG
912_RI	385013 > 385039	CTTTTGTGTAAGGAGTATGTACTAGC
WGAP129walk_1R	385855 > 385835	TCATCAGCACCAAAAACAACG
1174_RI	386308 > 386289	GAAAACAGCACAAGGCAACG
WTHIN1174_1R	387585 > 387611	CTTGTAATATGGTCGTTGTAATAATCGC
WTHIN1674_16F	388823 > 388796	CATATGTTAAGCTATATCTATGCACCAG



Primer name	Position	Primer sequence (5' > 3')
1174_LE	389080 > 389107	CTGGAGGTTTCTTTTATTGTACTATCTG
WTHIN1174_1F	389107 > 389080	CAGATAGTACAATAAAAGAAACCTCCAG
776_RI	389955 > 389926	GTAACCTAATAAATAAGTACTCTCTCAACG
WTHIN1674_15R	391106 > 391130	GGTGTGTTTCATTGTTTTGAGATAGC
776_LE	392597 > 392623	AGTAACTTGATATTTTGCAGTGTAGTC
WTHIN1674_15F	392663 > 392634	ACTACTCATTTTAGTACAACCTAAGTAGGTC
2808_RI	392802 > 392774	GTATGTTATGCAAACCATAAACTATTGAG
WTHIN2808_3R	393364 > 393388	AGATTGAATGAAAACAGTGTTTGGG
WTHIN2808_3F	393891 > 393872	ACCACCAACAACAGATACCC
WTHIN1674_14F	393933 > 393910	GAAGAGAGCATTTAACAACATCCA
WTHIN1674_13R	397364 > 397390	TGGTAGTTAGATATCATGCTGTTTAGG
WTHIN2808_2R	397365 > 397392	GGTAGTTAGATATCATGCTGTTTAGGAG
WTHIN2808_2F	398070 > 398045	CATGTTCTCCAACCTATCTTAGACACC
WTHIN1674_13F	398184 > 398159	TTACTTCAATACAAAGCTGATTTGGC
WTHIN1674_12R	399636 > 399658	AGAGTCAGACAACCTAAGCATGG
WTHIN2808_1R	399751 > 399775	CGTGTCTTAATGTACCAAGTTTCC
WTHIN1674_12F	400783 > 400763	CAAGTTGTCATAGCCTTGAGC
WTHIN2808_1F	400786 > 400764	CAACAAGTTGTCATAGCCTTGAG
2808_LE	403289 > 403311	CCGTGTTCTACGTTTAGTGTTC
WTHIN1674_11R	403995 > 404018	TCAAAAGTATCACATGTTTACACG
WTHIN1674_11F	404987 > 404964	GCATATTAGCTGATAAAGGAACCG
WGAP14_F	405892 > 405868	AGCCTAATTCTAGATCATCACACC
WGAP15_R	413104 > 413125	TGTTGGTCGATCAAAGACTCAG
WGAP14_R	413667 > 413640	TTTGATGTCTAGTTCATACATTTACCAG
WTHIN1674_10F	414450 > 414424	ACCTTATAAAATCAGCGAACTATAACG
WGAP13_F	417919 > 417945	TGCAGTTAACGATATTAGAATTGTTGG
WGAP13_R	418527 > 418501	TTACAAGACCTACTCTATTACTAACCC
WGAP12_F	425060 > 425086	AGGTGTTCAATAAGGTTGTAATACTG
WGAP12_R	425275 > 425250	GTACACTTTAGGACATATAACACTGC
WTHIN1674_9R	427161 > 427187	GATAGTATCTGTGTTGTAATATGCTG
WTHIN1674_9F	427951 > 427925	GTACCTACTTAAACATAACACTCATGC
WTHIN1674_8R	436561 > 436586	CTTCTGTAGCTCTTCTATAGTTCTCC
WTHIN1674_8F	437056 > 437035	AAGCAACAATGACGAAGTTACC
WTHIN1674_7R	462390 > 462413	ACAAGTATTTATTGCACACAGGTG
WTHIN1674_7F	463320 > 463302	CAAAGCGAGTAGGTGGAGG
WTHIN1674_6R	464416 > 464441	ACATTGGTATTGTCATACTTACTCCG
WTHIN1674_6F	465134 > 465112	AAGAACTGGAGTAGATGTAAGCG
WTHIN1674_5R	470438 > 470465	CTATGATAACAACCTTAGACATTTCTGGAC
WTHIN1674_5F	471498 > 471478	ACATCCATCTGCTGAACTACG
WGAP11_F	472684 > 472710	GCTATATCAGCTGATAAACTTGTTGAC
WGAP11_1640F	474112 > 474090	ACATATGCAGCTATGGATGTAGC
WGAP11_1640R	475854 > 475875	TGAGTGTGTTGTAGTGCAGAG
WGAP11_R	477196 > 477171	CAATAGCATTAGCTTTCTAAGTGGAC
WGAP10_F	480631 > 480658	GATGTTAATTTCTGGTTCTTATCTCTC
1458_LE	481320 > 481291	GCATATTACTTCCATAAAATCTTCACACTC
WGAP114_F	483116 > 483138	AGGAAAGTGCTTGTATTGTAGGG
WGAP114_R	483560 > 483533	CCGTATAGACTTAGTTCATAGATGAAACC
WTHIN1674_4R	483663 > 483686	TCTTCATCTACAGGTTTCAAGTATGG
WTHIN1674_4F	484973 > 484948	TCTATAAATAGCTCAGTACTGGAAGG
4518_RI	488477 > 488500	TGGATTAAGAAGACTAGCATCAGC
3205_RI	488808 > 488782	GTAGTGCTACTATAAAACCATTACCTG
WTHIN1674_3R	489262 > 489288	GTGATGCTACAAAATCATATACAGTGC
WTHIN1674_3F	490724 > 490704	TCGTCTTTCTAAGGAAGGCTC
WTHIN1674_2R	491999 > 492022	TGCACCTAAGATGTATAAAGTGCC
WTHIN1674_2F	493588 > 493562	CAAGATAAAGCTATACCTATTGAAGGC
3205_LE	493914 > 493939	GGATAGTTCATTAATTGATGGTCTGC
WTHIN21267_2R	494746 > 494770	CTGAGCTTGAAGATATTGTTTACCG
WGAP10_2178F	494793 > 494768	CATATAGCATTTCCTAATGCATCGG
WTHIN21267_2F	495472 > 495449	GCACTGCTAGTACTGATCTTAACC
WTHIN24706_33R	495771 > 495794	GGTATTGTTACGTACTTTTCAGGG



Primer name	Position	Primer sequence (5' > 3')
WTHIN24706_33F	496233 > 496207	ACTATACATAGCAATACTTACTGGCTG
WTHIN21267_1R	497454 > 497480	TCATTTAGTATAATCAGTGCATGATGG
WTHIN21267_1F	498293 > 498270	CATACAGTTTAAACGCTAACACTGC
WTHIN24706_32R	499133 > 499157	CGTAATTCTATAGAAGAAAGCAGGC
WGAP10_2178R	500006 > 500033	AAATTTATGTTCAAGATTTGTGTGCTC
WGAP10_R	500889 > 500860	CAACTATCATAGATAAAAATAACAGCTTTGG
WTHIN24706_31R	508762 > 508784	TCTAGTATTGGCATAGTGGTGTG
WTHIN24706_31F	509280 > 509259	CAAAGCAGTCACAAGATACACC
WTHIN24706_30R	515037 > 515059	GCTATTGGATGATTTGGAATGCG
WTHIN24706_29R	515301 > 515324	AATGCATAAACAGTTTCAGTAGGG
WGAP93_R	516221 > 516250	GAATATATGACCTCAGCTAATACTAATGTG
WGAP93_F	516699 > 516679	TCCTATTCCACCTGTCAATCC
WTHIN24706_28F	517019 > 516996	TCTATCAACAGAGGAACATCAAGC
WTHIN24706_27R	517288 > 517313	TTGAGTAGTGAGGTAATTTCTAGAGG
WTHIN24706_27F	518543 > 518517	CTCTAAGTAAAGTACTGAACTTTACAGC
WTHIN24706_26R	520785 > 520810	GTATGTATTCTTTACGTGAAGATGGG
WGAP29_R	522165 > 522191	ATTTTTATGCTTTACTACTGGAGATGC
WTHIN24706_26F	522276 > 522256	AGTTGAAGAGTGTATGGGACG
WTHIN24706_25R	523580 > 523605	TCGTAAGAATCATTAGTCAATTGGG
WGAP29_F	523671 > 52365	CACCAGAACGCCATCCT
WGAP30_R	524141 > 524158	TCCAGCACATGATCAGCG
WGAP30_F	524641 > 524620	CCACAAAATGCTGCAAAAATACC
WTHIN24706_24R	526080 > 526104	TCAATGAAATTAACACATCAGCTGC
WTHIN24706_24F	527864 > 527841	GGAGTAAAACATGCAACTTCTTGC
WGAP31_R	527897 > 527921	CATTTTTATCTCCTGAACCATACCG
WGAP31_F	528589 > 528560	ATTACCTCTTAAAAATTTACGAAACATCAC
WTHIN24706_23R	529007 > 529036	GTATCAAATTGTGTAATAGTTAAGCTTTCCG
WTHIN24706_23F	529990 > 529966	GCAACTCTTAATTGTGTAATCCGC
WGAP32_R	532303 > 532328	GTGAGATTTTAGCTAAGCATGATGAG
WGAP32_F	532681 > 532654	TAGTGATATCATTCAACACAATACACAG
WTHIN24706_22R	533631 > 533659	GATTATCTCAAATCTAGCTTCTCTATGTG
WTHIN24706_22F	535193 > 535164	TCAGATTACCATAAGTAATAATCTACCCAC
WTHIN24706_21R	545074 > 545099	TGTAAAAGTGAGAACAGAGTCTAACG
WTHIN24706_21F	545885 > 545869	TGGGGTTATGTGCTGGC
WGAP33_R	546541 > 546560	CTGCAAGAGATTGCGAAACC
WGAP33walk_2F	547678 > 547653	ACACACAACATATGTACAAAAGAGG
WTHIN_547962_R	547946 > 547922	CAAATAACAAAGACGATAGAGCACC
WTHIN24706_20R	548257 > 548279	GACAGCATTGTTTGTGTGATCG
WGAP33walk_1F	548446 > 548470	ACACAGTGGTAAACTATTGTCTAGC
WGAP33_F	549283 > 549264	AAAAACCAAGACCAACCTGC
WTHIN24706_20F	549599 > 549578	ACCAAGAAGAACACAAGACCAG
WGAP34_R	552572 > 552596	ATAAAAGTAGATTGTTGTGGGTAGC
WGAP34_F	552975 > 552947	CTACATTACCTTAATAACATTCACCACTC
WTHIN24706_19R	554517 > 554538	GGAAGGTTATGTGTAGCGATGG
WTHIN24706_18R	555548 > 555571	AGCGTGTAAATTTGATGCTTTGTGTC
WTHIN24706_19F	555572 > 555549	TGACAAAGCATCAAAATTAACACGC
WTHIN24706_18F	556494 > 556475	TGCATATAAGACGGCATGGG
WTHIN24706_17R	563249 > 563273	CTTGAAACAAATTGTCGTCTTCCTG
WTHIN24706_17F	564204 > 564184	GAATCCACGGAGTTTGAAGC
WTHIN24706_16R	566107 > 566130	AGGCGTGATACTATATTTTGAAGG
WTHIN24706_16F	567494 > 567475	AGCCTAACAGACAATCACGC
WTHIN22116_1R	567809 > 567836	GTATTCCTCCATTGATTATAACACGACC
WTHIN22116_1F	568135 > 568109	CATCAAATCCAAAAGTAGACTATACCC
WTHIN24706_15R	570298 > 570323	GTTTACGTTTGAATGCTCTTCAAGG
WTHIN_570833_F	570821 > 570845	AGAGTACACTATCATTCTGTTGGTG
WTHIN24706_15F	570846 > 570822	ACACCAACAGAATGATAGTGTACTC
WTHIN_572096_R	572084 > 572058	TGAATACATGAACCAAATTGAAGAACC
WTHIN24706_14R	572749 > 572771	GGTAATTTGCTGTATGTCTTCGC
WTHIN24706_14F	573401 > 573378	CACTTATTAATGTTCTCTGCCTG
WGAP35_R	574850 > 574875	TAAAGTGTATCGCAATATTAGTTGCC



Primer name	Position	Primer sequence (5' > 3')
WGAP35_13F	575449 > 575427	TGTATAACATCACCACCTTGCTCC
WGAP35_13R	577433 > 577460	TGTGTAATAATTGGTAATGCATTATTGG
16700_F	578018 > 577991	ACAGCTGTATTTACCATTTAAACATTCC
WTHIN24706_13R	580367 > 580391	TGTTTAAGGAATGACAAATCTCACC
WTHIN24706_13F	581133 > 581114	ACATATAGGCTGGTACAGCG
WTHIN24706_12R	582829 > 582851	CAGAACACGTGATACAAATTGCC
WTHIN24706_12F	583305 > 583281	GAACATCACTACCAATCTCTAAAGC
WTHIN24706_11R	585166 > 585186	CTGTCCTCCAGTTTCCATAGC
WTHIN24706_11F	585476 > 585458	AACACTCCACAACAGACCC
WTHIN24706_10R	586251 > 586273	TCAGTACGAAGTATTCTTGAGGC
WTHIN24706_10F	586830 > 586802	GTCATAAGAAGATATGGATTATCAGTAGC
WTHIN24706_9R	593237 > 593262	TGAACAAATGTCTTTAGTTGATGCTC
WTHIN24706_9F	594158 > 594133	GCATTTTGGTAAGGACATCTAATTCCG
WGAP36_R	595679 > 595706	TTAGTGTGAATGTGGTTATATAACAGTG
WGAP36_2F	596990 > 596968	ACATGGTCTCATAGAAGTTAGGC
WGAP36_F	597419 > 597399	AGAAACTTGCCCTATTCCAGC
WTHIN24706_8F	597747 > 597723	CACGATATGGAATTATGAGATGCAG
WTHIN24706_7R	601012 > 601041	AGCAAATTTCTATATCTAAGAATAACGTG
WTHIN24706_7F	602281 > 602254	TTAGTAATTTATCATAGAGCTAACACGC
WTHIN24706_6R	606101 > 606124	GTGGAAGTTTTTGTAGTGAGTACC
WTHIN24706_6F	607627 > 607603	AGTGTAGCTTCCATAGATAATCCAG
WTHIN24706_5R	608296 > 608316	AAGTACTGTGTGCAAGGTCTG
WTHIN24706_4R	608726 > 608754	GTTTCATCATTATGGAAATAGGAAAAGCTG
WGAP92_R	608978 > 608997	CTGCACAATTTGTTCCGGGTG
WTHIN24706_3R	609572 > 609594	TGGGTGTAGGGTTAATATTGGC
WGAP92_F	609881 > 609859	GCATCAAAGTAACTCTGCATGG
WTHIN24706_3F	610528 > 610502	GGTACAGATAGAAACCTAAACTGTAGG
WTHIN24706_2R	615392 > 615410	AGATCATGCAGGATGGCAC
WTHIN24706_2F	615900 > 615878	ACCGTTGTTATGTAGATCATCCC
WTHIN24706_1R	617979 > 618001	TCTTCGTGAGGTTCCATAATCCTC
WTHIN24706_1F	618250 > 618228	CCCAATCTAAACTTGCATAACCC
WGAP112_R	618458 > 618484	AGACTATATCACCATATCGTAAAGACG
1674_RI	619846 > 619820	CATTATGTCCATCATTAGATTCCAGCAG
WTHIN1674_1R	621066 > 621091	ACTTGCTATTCTTTATTCTAGGGGAG
WTHIN1674_1R	622066 > 622086	TGACTCACATGTTGCAGATGG
WTHIN1674_1F	622811 > 622785	GCCTTACACAACCTTATCATCATCTC
WTHIN1674_1F	622887 > 622865	TTCGGATCATTATCACCTGTGG
1674_LE	623274 > 623301	AATGGTAAAGATATTTCTAGTGTGGAC
1519_RI	624641 > 624613	ACTTAGATCCAGAGTTATTATATGACAGC
1519_LE	626131 > 626156	CAGTTTTTCATATGTTGGAAATGTGG
1309_RI	628639 > 628663	GATGTGAACATAAACCATTTTCCCC
2267_LE	629051 > 629028	TGCTGCTGATATAAATACTCCAGG
2267-2_LE	629326 > 629306	ACGATTATGTTGCACAGAGGG
WTHIN2267_1F	629465 > 629486	GCATAACAGCCAATAAGATCGC
WTHIN2267_1F	629596 > 629625	GCAACTATAAAAAAGTGATACACTATACGC
WTHIN2267_1R	630448 > 630420	AATGCTAGAGCATATAATAAATGTACCAC
WTHIN2267_1R	630672 > 630652	AGACAACGTATTGTCTGACCC
WTHIN2267_2F	632078 > 632101	GTAGTTTAAGATAAGCAGCGATCG
WTHIN2267_2F	632462 > 632493	GTATATCTGGAATTAAGATAGTGGTTAAGAG
WTHIN2267_2R	632919 > 632891	TGCTAATTTTACGCTATTTCAAATACACTC
WTHIN2267_2R	633283 > 633257	CTACTACTGTATCATCAGCCTTAGG
2267_RI	634044 > 634069	AGTTATTATCGTTGTCTATTGGAGGG
WGAP130walk_1R	634683 > 634704	TTGCACTAAATACAACAGCAGC
WGAP112_2F	636855 > 636828	ACAAAACCAATAACATACTGATAATGCC
WGAP112_1F	637533 > 637507	ACTAGATTACATAACTAGCCACTACAC
1309_LE	641215 > 641235	TCCAGATGACAATGCGTATGC
WTHIN24688_3R	646313 > 646334	CATGTGATACACCATTAGCTGC
WTHIN24688_3F	647161 > 647135	CCCTAAGTATAACAAATGTCTTACACG
WGAP91_R	649405 > 649431	AGCATTATATGTACATAAGCATCAACC
WGAP91_F	650879 > 650853	ACCTATAATAGCATACAGTAGGTTCCAC



Primer name	Position	Primer sequence (5' > 3')
WTHIN24688_2R	652228 > 652252	ATGTTTCTAAGTAACGACAGATTGC
WTHIN24688_2F	652853 > 652832	GTGTTTTTCGACGAAGAAACAGG
WTHIN_657239_F	657227 > 657249	CCAGAATGTGTTGCTACAATACC
WTHIN_657714_R	657702 > 657679	ACAGACTTATTAGCAAGTAGTGCC
WTHIN_660645_F	660633 > 660660	TGTTCTTAATCAATAATCATACCCTGG
WTHIN_661418_R	661406 > 661384	CCTCCACTTTGTAGTTTCTCAGC
WTHIN1360_1R	662830 > 662854	CAATTCTATACCATGTTTTCCACC
WTHIN1360_1F_C	663088 > 663063	AATGAACCACCATATTCAATAGATGC
WTHIN1360_2R_B	663751 > 663773	TGTAGCTAGCTGTACAATGAGAG
WTHIN1360_2F_B	664421 > 664397	ACCCAAACATTAATACTGAAGAACC
WTHIN1360_1R_A	665086 > 665112	AGTTTTGATGCATCATCTAATTCTGTC
WTHIN1360_1F	665357 > 665328	AAAGATACATTATGATTAGTAATCCTCTGC
WGAP90_R	666007 > 666032	GACTGTAATGAAATGTCATGACATGG
WGAP90_F	667254 > 667228	CATACACAACCGTAATTTACTACTCTC
WTHIN24688_1R	667872 > 667894	GTCTGTTAGTGAGAATTTCTGGC
WTHIN24688_1F	668421 > 668401	CCAATTAGCATCTTCACCTGC
WGAP89_R	669482 > 669502	TGCCTTCTGATCATGTTCCAG
WGAP89walk_1R	669799 > 669822	TGAAGATTCATGACACAATCTGC
WGAP89walk_2R	670220 > 670241	AGCATTCTGGTGTAAAATGCTG
WGAP89walk_2F	671093 > 671073	ACTTGCAAACCTCAACAACACC
568_LE	671677 > 671651	CTATGCAGCATAGTATTATCAGTTTCC
WGAP89_F	673249 > 673229	CAAGGTAAGAACAACAACGGG
WGAP88_R	679057 > 679082	GTTAAATCAGTTACAGTAGTACTGCC
WGAP88walk_1R	679774 > 679800	TGAGTAAGGGTAATTATTAATCGAGGG
WGAP88walk_1F	680482 > 680453	ACCTGCTTATAATTTATTATTGTCTACCTG
WGAP88_F	680960 > 680934	AGGCATTGAATCAAATATATGGATACG
WTHIN_681369_F	681357 > 681336	CTAGAACTGGCAGAATGTTTGC
WTHIN19567_1R	681478 > 681506	GTTCTAAAATTATAGCAGCAGTACTATCC
WTHIN19567_1F	681958 > 681935	GATGGTATATCTAGTTGGTGGAGC
WTHIN2007_1R	682328 > 682353	GACTGTAATGCTTAGCAATTGTAGAG
WTHIN2007_1F	682591 > 682565	CCCTTATACCTTATATTTGTACATGCC
WTHIN_684129_R	684135 > 684117	GTGCTTGGAACCAAGTGC
WTHIN_685216_F	685204 > 685226	AAAAGCTAAGGAAATCATGAGGC
WGAP87_R	686334 > 686357	GAGAGCAGCATATGTATGTTATGC
6654_F	687162 > 687134	GTACAACATTAGTAAGTTCTGAAAAAGTG
WTHIN568_1F	690695 > 690717	GTTTAAAGTTGCTAATGCTTGGC
WTHIN568_1R	691164 > 691140	AAGGAAATATAGGGTTAATGCAAGC
WTHIN_699689_F	699677 > 699702	GGAAGTTTACTTAGTCTGAAGTAGC
WTHIN_700846_R	700834 > 700814	AACATGCTTCATCTGTCTGCTGG
WTHIN_706368_F	706356 > 706379	TGTGCTTCTTATCATCTTACGACG
WTHIN568_2F	707467 > 707488	TCTAGTGGGAAGTACACTGGAG
WTHIN568_2R	708059 > 708039	GTCAGTTGAAGCAATAGCAGC
25104_RCF	709379 > 709401	AGTATTGATCATATCTGGCGGAC
699_RCR	710214 > 710188	GTATGCTGGTAAATCTTCTTATCCAC
699_RI	710687 > 710714	AATTACTACTGCTAAATAAAACGTAGCC
WGAP134walk_1R	711359 > 711384	TCTATAGCTGTATGCCTCATATATGC
WGAP134walk_1F	711965 > 711943	CAAACATGGCATGATAAACTTGC
10652_F	712830 > 712803	TGATAGTCAATTGTATAGTTAATGTGCG
WTHIN878_2R	713745 > 713769	GTAGAGAGAACTAGAAAAGGAGGTG
WTHIN878_1R	714794 > 714824	AGTATCTGATATGTGTTTTAGTATGGTAGAG
WTHIN878_2F	714829 > 714799	TGTATCTCTACCATACTAAAACACATATCAG
WTHIN878_1F	716011 > 715984	TGAACAATACAAAATCTTATACGACAGC
WTHIN_727336_F	727324 > 727349	CTAAAGTAACAGATAAAGAGCAGCTC
WTHIN2698_5R	727621 > 727642	TCTGAAATGGAGTTACGTGCTG
WTHIN2698_5F	729140 > 729117	CTATGAGTACAGAATGGATCAGGG
WTHIN2698_4R	730213 > 730238	AACGATTTAAGGTCTTGGTTAAACG
WTHIN2698_4F	731301 > 731278	ACTCATTTCTGAAGCAGGATAACC
WTHIN2698_3R	731560 > 731587	ACTGGTATTAATGATAGTCCATTACGTC
WTHIN2698_6F	731717 > 731694	ACTATACTGCAAAAATGAGTTCGC



Primer name	Position	Primer sequence (5' > 3')
WTHIN2698_3F	732756 > 732730	CAATGAATCGAATTTGTTGTTGTTAGC
10652_R	740548 > 740575	CAAATAAGCTATGTTTTCTTATAGCTCC
WGAP85_F	740732 > 740712	AACACCAAGCGAAGAAATTGC
WTHIN2698_2F	741862 > 741838	CACACAGTCTTTATATGTTGAAGGC
WGAP84_R	743171 > 743194	CATGTATACCTAATGCATGATGCC
WGAP84_F	743490 > 743463	ACATAGTGATACAGAAGGTATAAECTAG
878_LE	744134 > 744162	CATTACTAAAGTACCTCTCTTAAAATCGC
687_LE	744928 > 744903	CAGAATCCTTAGAATCTTATGATGCC
WTHIN282_3R	745378 > 745397	GCACTCTTGTAACCAACAGC
WTHIN282_3F	745944 > 745919	CATACTACTTCAACTTCAGGTAGCTC
WTHIN2698_1R	746426 > 746448	TCTACAGAATCCCTGTTTATGGC
WTHIN2698_1F	747184 > 747162	TCAAGTGTACCAACTTTTCCTCC
WTHIN687_1F	748247 > 748275	TTAGAAGATAATTCTGGAACTAAAGAAGC
WTHIN687_1R	749165 > 749141	TGGAGCTAGATAATCTGGTAATGG
WTHIN687_2F	750693 > 750717	CTCAAAGGAATACAAGACACTATGC
WTHIN687_2R	751338 > 751313	CAGGTAACGTTTTACACATATTGGAG
687_RI	751583 > 751604	AGATCATCACGACTCATACCAG
3395_RI	752565 > 752539	ATCCATAGATCAACTTGTATTAGCG
3395_LE	756698 > 756723	TGAAGATGATATAGATACACCAGCAG
WTHIN8190_3R	757306 > 757324	CTGAACGTGCTTGGTTGTC
WGAP108_F	757331 > 757311	TACAATTGACAACCAAGCAGC
WTHIN8190_3F	758853 > 758821	CACAATAAAAGATAACATACTGTTTAAATCAC
WTHIN8190_2R	763133 > 763156	ATCCAAAGCCATTAATTCCAAGTG
WTHIN8190_2F	764033 > 764004	GCATATGAAATACTGAATACGAATGATAGG
WTHIN8190_1R	764505 > 764525	AACCCAAGTGGTGATAGTCTG
WGAP111_R	765630 > 765597	CATTCTTTATAATTAGGTAAGGATATGTATCAC
2546_LE	766075 > 766049	TTGTTCTATTTTAGATAAACCACCACC
WTHIN_773698_F	773686 > 773707	AGAGAGTACCAGAAACTGAAGC
2546_RI	773844 > 773863	ACAAGTACAGTGTCTGTACG
WGAP135walk_1F	774156 > 774181	GAAGCATTAAATAGCTGAAGATGAAGC
761_LE	775710 > 775685	CTTTGATCTGGTAAGGATAGTAATGC
761_RI	776335 > 776361	CACAATAATAGACATGTTTGAAACAGC
WGAP135walk_1R	776980 > 776957	TGTTTTATCCTGTAGTTGTGTTGC
WGAP111_F	777769 > 777744	GCAGTTTATAGCATCTATTAGGTCAC
WGAP110_R	781450 > 781476	CTACACACATAACAACAATAGTTAGCC
WGAP110walk_1R	782077 > 782106	CCAAATTACTACATCAAAATACCTAACCCAC
WGAP110walk_1F	782818 > 782789	TGTAGCTCTATATAATGTTCTGTATGTGG
WGAP110_F	783197 > 783178	AGTTGGAGTCAAGATGAGCG
WGAP109_R	791387 > 791416	AAGTCATAATGCTATGATCTTAAGTATGTG
1932_LE	792365 > 792344	TGCTGTTTGAGGTAAAGTCTGG
1932_2RI	795195 > 795219	TGTAATTTCTCTATACGACCGATGC
1932_RI	795400 > 795427	CATAGAAATCTGATACCATCTTACTTCG
WGAP109_F	796803 > 796778	AGCAATTGAGAGAATTTTATGTCAG
WGAP108_R	799303 > 799331	TCAATAGATTCAATACTCACTGCTATAGG
WTHIN8190_1F	800281 > 800251	TCTTTATAATTAGGCAAAGGATATGTATCAC
WTHIN8190_4F	800611 > 800587	CAGTAGTAATCAAGCAGTAACTGTG
WGAP83_R	804671 > 804691	ACCTACACGAAATACTTCGCC
WGAP83_F	804977 > 804950	GCTTATTACATTAACAGAAGCTAAGCAG
WTHIN12708_7F	805423 > 805401	TGATAGGATGGTTTGCTACTAGG
WTHIN_808166_F	808154 > 808173	TCCTGTGGTAACTTAGCAGC
WTHIN_808749_R	808737 > 808717	TGCTGCTGTATTAACCTCCACC
WGAP82_R	809362 > 809392	GTAGATTTTCATTATACAAATACCTTTGCAC
WGAP82_F	809740 > 809719	CTGTGCAAGTTACTTCATGTCC
WTHIN_811768_F	811756 > 811781	GAGTAACTTTGTGTGTAGTATCTTGC
WTHIN_813027_R	813015 > 812992	TTTCTAAGAAGAATGTGAAGCAGC
WGAP81_R	813322 > 813343	TGCTTTCAAAGATTACCACC
WGAP81_F	813899 > 813877	GCTAATGAAATTGCAGTTGATGC
WGAP80_R	819878 > 819903	AAGATAACATGACAAATACTTGCCC
WTHIN_820343_F	820329 > 820351	TCTGTAACCTAAGTCAGCTGTTGC
WGAP80_F	821189 > 821164	AAAGGATTAACACTAAATAAGGTGGG



Primer name	Position	Primer sequence (5' > 3')
WTHIN282_4R	829220 > 829240	GTTTGTAAAGGGTGCTACTGC
WGAP79_R	829676 > 829708	CAGTGCCTTATTGTAATATATACAACATTTCTC
WGAP79_F	830701 > 830677	AATTTGGATGCTAATAACGTAGCAG
WGAP78_R	837879 > 837904	AATTTGATAATGTGAGATGGCAAGAC
WGAP78_F	838322 > 838297	CAAATCCAAGTTCAGAAAAGTAGTACC
WTHIN12708_6R	844241 > 844264	CTGGTATATCATGTTTTGGTGGTG
WGAP77_R	845088 > 845116	TAACCTACGAATTAATTTCAAGAACCAG
16700_R	845673 > 845651	ATGGTGTAACACAGCAATAGGAG
WTHIN12708_6F	846066 > 846046	TGGCAAATGTAGTACCAGGTG
WTHIN12708_5R	847499 > 847520	TGCATCTAATGTAGGAGCATCC
WTHIN12708_5F	848643 > 848618	ACTTGAGTAAGGGTATTTGTTAAAGC
WTHIN12708_4R	849226 > 849251	GCATAATATGGTATATCACGTTCCAGG
WTHIN12708_4F	850269 > 850248	CAGCGTCTTACGGATATCTAGG
WTHIN12708_3R	854446 > 854470	AGAATCTGTGCCATCTATATACTGC
WTHIN12708_3F	854970 > 854948	GATGAATTTTCATCCTGCATCTCC
WGAP28_F	867655 > 867679	ACCACAATATCTCCAGATGATACTC
WGAP28_R	867882 > 867860	TGGAATTGATGCTGATACTGGAG
WGAP27_F	868593 > 868617	ACCAATCCACAAAATACTAACTCAC
WGAP27_R	869125 > 869096	TGATTAGTCATAATTTATGGTAGAACTTGG
WTHIN12708_2F	869785 > 869760	CCAGCATTGTAATTATCTTGAAGTGC
WTHIN282_2R	870812 > 870838	AAGTATAATACACACAATAACTCTGCC
WGAP26_F	870963 > 870994	ATTCATTAATGGATAAAGTACTTTATCAACC
WGAP26_R	871934 > 871905	GAATTTTGCATTTAGAATGATTGTATTGG
WTHIN3611_1F	875278 > 875304	CAATTAACAAGCAGCAAATTATCCAC
WTHIN3611_1R	875739 > 875708	GTATCAAAGTGATAAAAACTTTAGTATGGTAG
WGAP25_F	876759 > 876785	AAAGCGTAACATTTAATGTACCTAGAC
4015_RI	877831 > 877807	TTTGCTAACACATTTTACCAGTAGC
WTHIN656_2R	881655 > 881674	TCCACATCTGCACTATCACC
WTHIN656_2F	883235 > 883212	AGTATTAGAACCTGAGTTACAGGC
WTHIN4015_1R	883583 > 883609	TCCTAATAACACATTTTTCATCACTAGC
4015_LE	884836 > 884856	GCATACTTGATTGTGCAGCAG
WTHIN4015_1F	884856 > 884836	CTGCTGCACAATCAAGTATGC
WGAP25_R	885290 > 885270	TGTTGCTAAAGGCTATGCTGG
WTHIN12708_1R	887946 > 887968	ACTCTATCAACTGCCAAGATAGC
WTHIN12708_1F	889495 > 889472	GGATGATGGATCAACTAAGACTCG
WGAP24_F	902868 > 902891	AACTCTAGATCTCCTGAAAGTAGC
WGAP24walk_1F	903819 > 903846	TTTATCGGTATTACAACACTCATTAC
WTHIN656_1F	903849 > 903824	TTTGTGAATGAGTAGTTGTAATACCG
WGAP24walk_1R	905594 > 905570	AGGAATTTAGCACATATGATTGTGC
WGAP24_R	906184 > 906155	TTTGTAGGTATACATTTGAGTTGATATTGG
WTHIN_907406_F	907392 > 907418	AGAGAGTCAGTATACTACATAACAACC
WTHIN_908473_R	908459 > 908433	CAAAGTATGATGGGTTATATGTCATCG
WGAP23_F	918144 > 918173	TCCCAGATACATTTAATAATGATGATTAC
WGAP23_R	919004 > 918978	TTACAAGGTAGAGTTTTGTTAATACGG
WTHIN13950_1R	921036 > 921058	TGCTCAAAACTCAGCAATTACAG
24983_RCF	922374 > 922344	TTTCAATATAATAAGTGTATTAGGTTGAGTC
WTHIN13950_1F	922711 > 922684	AACTGATGTTACTATTTTAGGATAGCAC
WTHIN282_1R	923405 > 923430	TGTGCTATACTAGTTCAATAGCATCC
WTHIN282_1F	924122 > 924101	AGTCTTCCAACAACACTTACCG
WGAP22_2R	925773 > 925797	AATACCTACAATCAATCTAGACCCG
WGAP22_F	925921 > 925948	AATACTGATTTACAGCAGATTAGCAG
WGAP22walk_1F	926342 > 926369	AAACCCTATTAATAGTAAACCTACTAGC
WGAP22walk_2F	926909 > 926936	GTAAGGAAGAAAACCTTATCCACTATCG
WGAP22walk_3F	927530 > 927549	TCAACTGCCTCTGAATGTGC
WGAP22walk_4F	928127 > 928148	GGGTATGTAGGAAACTTGGTGG
WGAP22walk_4R	928760 > 928740	TGCTTTGTGTTATCGCCTACG
WGAP22walk_3R	929486 > 929462	TGGATGTATAACTTAGGTGCTGC
WGAP22walk_1R	930111 > 930085	GATGTATTCACATGTTATCATGAGAGG
WGAP22_R	930592 > 930567	TGAACCTGGGTATTATAAGATGGAG
WGAP22_2F	930700 > 930681	GCATGGTACAGGTCATGGTG



Primer name	Position	Primer sequence (5' > 3')
WTHIN568_4R	931810 > 931830	AAGATCCAACCTCCTGTTGAGC
WTHIN568_4F	933035 > 933008	AACCTAATATTCGTTCCATACATAGCTG
WGAP21_F	936478 > 936498	TAGTTGCCCAACTATAGCAC
WGAP21_R	937548 > 937522	AAGGTTTATGGACATTTAGCTTATAGG
WTHIN_940940_F	940928 > 940953	CTACCGATAACTGAAAAAATAACCC
WTHIN_942230_R	942218 > 942197	TCCGCTTTTTCTGTATATATCCC
WGAP20_F	943418 > 943441	TGTAACATTATCTCGCTGCATAGG
WGAP20_R	944607 > 944582	ATTATGTAACCTGTTGCGTATTGTCAG
699_RCF	954033 > 954060	CATCAACCTATGTCTAGTGAGTATATCC
WTHIN568_3R	954533 > 954556	ACACCAAGATCTCATTGATAACG
WTHIN10217_1R	955428 > 955454	TCAATTTTCAAATACCCAATAACTCCC
WTHIN568_3F	955888 > 955864	ACTCACGTATAACTGATCTATGCAG
WTHIN10217_1F	958420 > 958386	TGATATTAACCTATCTATCTTTTGTAAAGTGTACTG
WGAP76_R	960661 > 960687	CACAATCATATGATTTTGAAACTCTGC
WGAP76walk_1R	960942 > 960970	AGCTAACAGTTTACTGAATAACATATCAC
WGAP76walk_1F	962014 > 961989	AGAACTCTAAATGATAATGTCTCGGG
6654_R	963134 > 963109	TGATGAGTAGATTACAAATAGTTCGG
WTHIN11030_3R	966121 > 966145	AAGTATATCGCAACATAATGTGAGC
25104_RCR	966217 > 966238	TTTCATAGCATCCTACCTTCCC
WGAP76_F	966242 > 966224	TAGGGGGAAGGTAGGATGC
11030_RI	966520 > 966492	GGTTTTGAGAGTAGTAGATATTTGTTAGG
WTHIN11030_3F	966660 > 966640	TGACCGAAGTGAATACCAAGC
WTHIN11030_2R	966978 > 967001	AGTTCGCATTCTTTAGTTTTGCC
WTHIN11030_2F	968032 > 968008	GTGAACATTTACTAGTGGTACATGC
WTHIN14308_2R	968270 > 968299	TCTCTAGAAAAGAGAATTACTGTTAGATCC
WTHIN14308_2F	968620 > 968597	ACAGTAAAGTTTGTAGAGGATGGG
WTHIN11030_1R	972270 > 972293	TTCACATTGAGTTACAATTCAGC
WTHIN11030_1F	973098 > 973072	TCAGTTATAGAGTAAAGAGTTAGGACC
WTHIN_978928_F	978916 > 978943	CAATCTTAAGGTGATTATCTGTATAGCC
WTHIN_980451_R	980439 > 980414	AACAGTTATACTTCAAGATGACATGC
WTHIN14308_1R	980658 > 980686	GATAGATTATTCTTGAAGATCCATTGAG
WTHIN14308_1F	981332 > 981307	GGGTGAACAAGAATTGAAATTATTGG
WGAP75_R	981789 > 981815	AGTCAATCATTATCAACCAACAAATCC
WTHIN12427_3R	982419 > 982441	CTCAATCATGTATCCGTCACAC
WGAP75_F	982426 > 982398	TGATTGAGATATAGATGCTATAGTTGTCC
WTHIN12427_1R	982729 > 982757	CATGCTACTTTTACCCTATATACAAGAC
WTHIN12427_2R	983518 > 983543	CCAAAATACAATCTACGTATTCTCGC
WTHIN12427_2F	983744 > 983719	GTTGGTGCAGTAATAGTGTATAATGG
WTHIN12427_1F	984446 > 984416	TTATGACTTATGAGTCATATATTAAGCCAG
WGAP74_R	1000142 > 1000163	TCTCCATAGCAGCATTATGCAC
WGAP74walk_1R	1000674 > 1000698	CTCCAGCTTCCTTAATTTTGATAGC
WGAP74walk_2R	1001178 > 1001204	GGAAGTACAGTTTTTATGGAATGTAGC
WTHIN_1001515_F	1001545 > 1001571	CAGAATTACAACGCTATATAGTAGCAG
WGAP74walk_3R	1001916 > 1001936	TCATTCCTGTGTTTCACATGC
WGAP74walk_4R	1002575 > 1002601	CATTATGTCCTTAGTACTCATAAGCAC
WTHIN_1002866_F	1002852 > 1002879	CTCACTTCAATCATCTAACAATATAGCC
WGAP74_F	1004365 > 1004344	GGGAAGTTTGTATCCACAGG
WGAP73_R	1005208 > 1005234	CTATAGCACGAAAATTATTATGACGCC
WTHIN_1005552_F	1005538 > 1005561	CCAATGCCTAATACTACATTAGCAG
19566_F	1006580 > 1006555	TTGTAATAGAAAAAGTGGCTTTGGTG
WTHIN2177_1R	1021310 > 1021331	AAGCATGAGTCCTAATACTAGGG
WTHIN2177_1F	1022226 > 1022204	AGTTACGTTGTTGCATTTAAGGG
2177_RCR	1022347 > 1022370	AGGTAACGTGAAGCTCATTAGTGC
2177_RCF	1023930 > 1023907	GGGTATAACTCAATGGTAGAGTGC
19566_R	1028526 > 1028549	CAGAAATTCAGCAGAGTTTAGACC
WGAP72_F	1031212 > 1031189	GGAGTGTAAGTATGGGTATAGCAG
WGAP71_R	1034459 > 1034484	CCATATTACTTCATCAAACACCTTGC
WGAP71walk_2F	1035673 > 1035648	ACAGTTCAAATGATTACCTTATGACG
WGAP71walk_1F	1036318 > 1036293	GAATTATGGTTTGTGTTATTGCATGC
WTHIN1575_1R	1036441 > 1036466	ACACTTAATAAAATCAACATCTGCC



Primer name	Position	Primer sequence (5' > 3')
WGAP71_F	1037712 > 1037686	ATTGATGTGGAATATTGAGTACATAGC
WGAP70_R	1041036 > 1041062	TTCTCCTGTAATATCATCACTTAGCTC
WGAP70_F	1042166 > 1042139	GAGACAAAAGTATATTTTGTAGCCTGTG
WTHIN1403_1R	1043355 > 1043379	AAACAATCATTGATCTACAGCAC
WTHIN1403_1F	1044372 > 1044350	AGAAGTCATGAGTATCGTTACGG
WGAP69_3R	1044428 > 1044453	ACTATATGTAGCAATGACTAGAGCTG
WGAP69_2R	1044786 > 1044811	TTTATACAGTACAATTAATGCCGCTC
WGAP69_R	1044888 > 1044916	CTGTACCATAACTAGCTACATATAAGACC
WGAP69_F	1045897 > 1045872	TTTTGATTTGTCTATGACAGCTATGG
WTHIN1486_8R	1047049 > 1047076	ACTTAAACTCTGTTATCAGATTACTTGC
WTHIN1486_8F	1047916 > 1047892	AGTATATGATAGGTCAACCTTTGGG
WGAP68_R	1054385 > 1054406	ATAGTCATAGGACAAGGATGCC
WGAP68_F	1054602 > 1054580	ACTGTTGGAAGTATAGCACTTGG
WTHIN1486_7R	1057517 > 1057541	CAAGGAGTTCAAGTAATACCTATGC
WTHIN1486_7F	1059139 > 1059110	TTTTACTTGTCTTAGTGTATAATATCCCG
WTHIN1486_6R	1062578 > 1062603	AACATCAATAATCAAGAAGTCATGGG
WTHIN1486_6F	1063428 > 1063403	TGTACATGTGATATATTGAGGGATGG
WGAP67_R	1064530 > 1064552	GCATGCCCTGATATTTGAGAATCC
WGAP67_F	1064981 > 1064954	CAATGTTGCAAGTAAATCTATTAATGGC
WTHIN1486_5R	1065182 > 1065206	CTATAGGCTTGAAAATCACAAGACC
WTHIN1486_5F	1066325 > 1066306	TGTGACAAGACTGGTACAGG
WTHIN1486_4R	1067000 > 1067027	AACTTCAAATAACATTCATGGTATAGCG
WTHIN1486_4F	1067825 > 1067803	GAGGTGGATTTAATCTGTCTTGC
WTHIN10971_1R	1071558 > 1071585	CACTCTTATAGGTATAATACAACGTTGC
WTHIN10971_1F	1072938 > 1072913	CTTACGTTAATAGAAAAGGTGGTGTG
WTHIN1486_3R	1073040 > 1073065	ACATATCAGCTTCTAGTAAAACCACC
WTHIN1486_3F	1074680 > 1074658	TGACAGGAGAATTTATTGGGTGG
WGAP66_R	1077357 > 1077375	TGCAACAGCTGAAGTTTGC
WGAP66walk_1R	1077985 > 1078005	ACGCAATATTAGCACCTGTCCG
WGAP66walk_1F	1078306 > 1078286	GCTTTGCTTGATGCATATGGG
WGAP66_F	1079440 > 1079415	TGATATGTATGAATCGTTTTTAGCCG
WGAP65_R	1085361 > 1085389	GAACTACCATTGTTCCCTTATATTATCACC
19572_R	1085553 > 1085527	CTAGTTGTTACAACAATTTTGATTGGG
WGAP65_F	1086154 > 1086133	TAGCAGAAAAATTCATGGTGGC
WTHIN_1086186_R	1086175 > 1086151	AAGATCCTATGTGAGAAACACTAGC
19572_F	1110043 > 1110067	CTTACAAGAAAGGCATAAATACCTC
WGAP64walk_1R	1110432 > 1110458	ACTAATATGCTAAAACACACTTATGGG
WGAP64walk_1F	1111606 > 1111584	GGTGCCATCTTTGTTAGTAATGG
WGAP64_F	1111768 > 1111741	CCATGTATATAGTGTATCTTTGAATGC
WTHIN1486_2R	1114274 > 1114299	CTGTTGCTACTGTTATAACTAAAGGC
758_RC1_F	1114712 > 1114732	AAGCAGTCAAATACCCATCCC
758_RC1_R	1115452 > 1115429	TCATGATGTAGTGAGAAAGTGTGC
WTHIN1486_1R	1115557 > 1115582	TCCTGACATAATAGATTCTATCACCC
WTHIN1486_2F	1115583 > 1115558	AGGGTGATAGAATCTATTATGTCAGG
WTHIN1486_1F	1116378 > 1116352	GTGTAATCATGTATTAAGTGATGAAGC
WTHIN_1125301_F	1125168 > 1125191	TCTCCATCAGAAATAGATAACGAC
WTHIN24731_6F	1125168 > 1125192	TCTCCATCAGAAATAGATAACGACC
WTHIN_1126360_R	1126435 > 1126410	GTATTGCTACGGTAATTTATCTGGTG
WTHIN_1130794_F	1130869 > 1130888	TGTTAGGACATCGTGGATGC
WTHIN_1131423_R	1131498 > 1131474	CCTTATTGAATATTGTGCATTGTGC
WGAP63_R	1134132 > 1134162	AGTATAGATTCTACATTCTACACTACACC
17139_R	1134211 > 1134239	AAAATAGTCGTAACACTAGTAGCTTATCAC
WGAP63_1462R	1135280 > 1135252	TTACGATTAATAATGGTGATGTATGGG
WTHIN_1136833_F	1136908 > 1136929	TTACTGGTTCAGTATGAGGCTC
WGAP63_1462F	1137707 > 1137732	CCTTGATACATCTTTTATATAAGCACC
WTHIN_1137750_F	1137825 > 1137804	GCAGGTCACCTTATACAGTCTGG
1547_LE	1139197 > 1139169	ACTTTATAGCTAGTAGAGTGTGATATTGC
1547_RI	1141061 > 1141087	ACCAACAAAGAAATAGTTAACACTACC
896_LE	1141633 > 1141605	TTTAAGGTACATCTATTGTTAAGGAGTG
WGAP63_F	1142673 > 1142648	CATAGTATTAGCGGTAGTACGTACAG



Primer name	Position	Primer sequence (5' > 3')
WTHIN13621_1R	1142854 > 1142883	TCCTTTAGATATTCAGTTACAACATGTAC
WTHIN13621_1F	1144089 > 1144064	CCATAGATTCTAGTATCAGGCAATGC
WGAP62_R	1147598 > 1147623	ACACATAATTACACATCATTGAGCG
WGAP62walk_1R	1148141 > 1148168	CTTACACTCTATAGTATATCCCGTTAGC
WTHIN24165_2R	1148782 > 1148806	AACATATTGTATGACAAAAGCTCTGG
WGAP62walk_2R	1149141 > 1149169	TGTAATATCATACATATACTAACAGCGGG
WTHIN_1149078_F	1149185 > 1149211	TTTCACAATACTCAGTCACTTAATACC
WTHIN_1150475_R	1150550 > 1150525	CATATACCTTGTCTTAAGGTAACGG
14311_R	1150656 > 1150634	AAACGATTTTTGGTGCTATTGGC
WTHIN24165_1R	1163877 > 1163894	CATCATCAGCTGCTTGCC
WTHIN24165_1F	1164141 > 1164114	GTTGGTATGTTATCAGTTATTCTCAAGG
14311_F	1170927 > 1170956	TCAAGGTAAACCAATACTACATATTACCTC
WGAP61walk_2R	1171088 > 1171112	ACATATCCACACTTACTAAACCCTC
WGAP61walk_2F	1172236 > 1172212	GTTATTAAGAGCATAGGCAAGTAGC
WGAP61walk_1F	1172343 > 1172318	ACTGGTATAATGATTATGAATGGGGG
WGAP61walk_3F	1172389 > 1172365	TGGTAACTCTATTTTGAAAGATGCC
WGAP61_F	1172904 > 1172878	ATGTAGGTAAATAGCTAAGGTGTATGG
WTHIN7701_1R	1174702 > 1174723	GCTTACGAACTGTAGAATTGGC
WTHIN7701_2R	1175366 > 1175390	TGGAGAATTATCAAGTATGTCACGG
WTHIN7701_1F	1176337 > 1176315	TGATGCAGGTATCATATTGGTGG
7702_LE	1186937 > 1186960	TTCTACTAGAGTGAGGGTTTATGC
5054_RI	1187351 > 1187328	TGAAATTTCTAGGGTCAGGATGTG
WTHIN5054_1R	1193363 > 1193389	TGATTACTATACAATAACGTGTGGTG
WTHIN5054_1F	1194065 > 1194042	TGCATATTTCTGCAATGTGTACAG
5054_LE	1195073 > 1195101	ACACATAGATAAAGTTTTTGACACTATGC
WGAP121walk_1F	1195326 > 1195351	CAGCTAAAATAATTTAGCTTTGCACC
WGAP121walk_1R	1196279 > 1196246	TCTTAGATAGTTAATCAGTCTTACAATAATCC
7239_RI	1196509 > 1196489	GTTGCTCCAAAAGAGTTCCAG
WTHIN1674_2R	1201095 > 1201121	CTTAACAGTAGTTTCACAACACTAGAAGC
WTHIN1674_2F	1202431 > 1202409	CTGACGTTCAAACACTATAGAACGC
7239_LE	1206918 > 1206940	GCATGATGCCTTCAACTAATTCC
3850_RI	1207731 > 1207703	GTTACACACCATTACATATCATATAGAGC
WTHIN896_1F	1210444 > 1210467	TCTCACTATTCAATGGTCTATGGC
WTHIN896_1R	1210802 > 1210777	AAGAGAATATGCATAGATTGAGGGTG
WTHIN896_2F	1212816 > 1212843	AACCTATATATTCCATTCAATGCTTAGC
WTHIN896_2R	1213483 > 1213460	TGGTGAAGAAGATATTCAGTTGGG
3850_LE	1214106 > 1214134	ACCTATATTTTGTGAATCAGAAACACAC
WGAP188walk_1R	1214639 > 1214671	TCCTTTAGAGATCTATATTTATAATTCAGCAC
8192_LE	1215853 > 1215828	TGTGTCTTATGAATTACATGATGCTC
8192_RI	1226711 > 1226739	ACAATTTCACAATATCTTAATACCAGCAC
4307_RI	1227652 > 1227631	GGAATACAAGTGATGATGGTGC
WGAP107_R	1231739 > 1231766	CTTACCAATTTCAAGTCTTAGTATGTCG
22799_R	1232239 > 1232213	TGTGATTGTGTTTGAATACTAATGGC
WTHIN15591_14R	1233931 > 1233954	AACTTAAACACAATTACTGTGCCG
WTHIN15591_14F	1235634 > 1235611	GCCATTCCCTAACTATGTACAGTG
WTHIN15591_13R	1240369 > 1240396	CAGCAATCATCAGATATATACTTCACAC
WTHIN15591_13F	1241467 > 1241441	ACAGTACCAATATGCTATTAAGGTGC
WPCRJL1_check_1F	1259546 > 1259572	AACACACTGCTCATTATATATACATGC
WTHIN15591_12R	1259630 > 1259656	TCACATTAACATCAAAGAATTAGGCAG
WPCRJL1_check_1R	1260938 > 1260912	AGCACTACTAACAAATAACAAATACCC
WTHIN15591_12F	1261306 > 1261278	AATAACACAACATAAAGGTTAAAGTCTCG
WL2AP1_check_1F	1281261 > 1281282	ACATTCCCACTTAACTGCATGC
WL2AP1_check_1R	1282759 > 1282730	GGTATGTTTATCAGTTGTTAAGTACAAAGG
WTHIN15591_11R	1290476 > 1290499	ACAGTAAACAGCAAATATGTGTAGC
WTHIN15591_11F	1292019 > 1291994	TTTTCGATAACTTTAGAATTGGGAGG
WTHIN15591_10R	1293749 > 1293775	CACTCATATATATCATGACATCAACGG
WTHIN15591_10F	1294969 > 1294946	CTCAGTCTTGTAATTTGTCAGCG
WTHIN15591_9R	1295571 > 1295598	AGCAATTCATATTTATGCTACAGTATCC
22799_F	1296151 > 1296177	TTCCAACCTTTCTATCTGATTTTGCAG
WGAP48walk_1R	1296952 > 1296979	CACAAAACCTAGGACTGTTAAAGTTAGAC



Primer name	Position	Primer sequence (5' > 3')
WTHIN15591_9F	1297159 > 1297132	TCATTAAGTGAAAGCTATTGTAATGCTC
WTHIN15591_8R	1297532 > 1297562	ACTACTACTACTTCAATATCAGTTAATCCAC
WGAP48walk_1F	1298437 > 1298408	TCAGATTATTAGTCGTAATATTATTGCTGC
WTHIN15591_8F	1299127 > 1299107	AGGCATTATCAAGAGGTGCAG
WGAP48_F	1299349 > 1299374	AAGTAGACAATAAAATACTCGCTTTGTC
WTHIN15591_15F	1299493 > 1299514	TCGTAACTTCCACAATACTCCC
WTHIN15591_15R	1300267 > 1300245	AGTGAACAAGAAATCTTGGATGC
WTHIN15591_7R	1300492 > 1300514	ACAACCCTATTGTACGATTACGC
WTHIN15591_6R	1301366 > 1301389	AATAACGTGCTTTTGGTCTAATCC
WTHIN15591_6F	1302260 > 1302235	TCATAATTTGTGGTTGAAAACGAGAG
WTHIN15591_5R	1306121 > 1306147	GATAACCCACTTAACCTGTATAATCAC
WTHIN15591_5F	1307235 > 1307209	GTGACATTAATGACATCAACCATAACC
WGAP47_R	1307594 > 1307623	TGAATCTAGTAACATATGTGATTTGTACAG
WGAP47_F	1308520 > 1308498	TTTGGTATTGTTGTGTGAAGCAG
WTHIN15591_4F	1308727 > 1308707	GTCTTTGTCACGACAAATGCC
WTHIN15591_3R	1312053 > 1312078	TTGTACAGTAGATCTGTACAATACCG
WTHIN15591_3F	1312865 > 1312842	GATCTTCTTGTGAGTTACTTGGG
8028_RCR	1313586 > 1313606	TTGTATTGCCAGTTGTTGCAC
8028_RCF	1316525 > 1316499	ATGTGTGGATCTAGTAATTCATTAGTG
WGAP46_R	1320481 > 1320504	AGTAGACAGTATAACAAGCGTTTCC
WGAP46_F	1320949 > 1320922	GGTAAGAATTTTGTAGTGAATTTGTTGTC
WTHIN15591_2R	1325023 > 1325042	TGTTGACACACTGACATCC
WGAP45_R	1325281 > 1325309	CAGCAACTATCAAATATACCAAATTGAC
WGAP45_F	1326066 > 1326046	GAGTCTGAAAACCCATTTGCC
WTHIN24586_1F	1330242 > 1330269	CTAAATACTGAGAATTAGGGAACAAACG
WTHIN24586_1R	1330865 > 1330839	TTGTGAATTCTAGTATCAATCTTGTGG
WTHIN15591_1R	1331315 > 1331335	AGTAAAAGCAGATGGACTCGC
WTHIN15591_1F	1332066 > 1332044	CTAGATTTTCACCTTGTGCTACG
WGAP44_R	1334531 > 1334557	TGTATCAAATACAATTAGTAGCACCAC
WGAP104_F	1335112 > 1335086	CATGCTAATCTATGTGACAGTAAAGTG
WTHIN15591_17R	1336756 > 1336777	TCTGACTTAGCAGCAGATAACC
WTHIN15591_17F	1337040 > 1337020	ACAGGTATTCAATGGTGGAGC
WGAP43_F	1343829 > 1343852	GCAGATGATTATGTAACAAAGCCC
WGAP43_R	1344151 > 1344131	AGCTTTTTGCGCAACTTACAC
WGAP44_F	1347063 > 1347087	CATGCATTACACAGATCTTCAACAG
WTHIN2266_1F	1348121 > 1348096	CCACACTTTCAGTTCAATCTTATAGC
WGAP104_R	1348480 > 1348451	GAAATAGGTGTTTATTTCTGTTAGAAATGC
WGAP105_F	1352191 > 1352214	ACACGTTACCTACTCGTAAAACAC
WGAP105_R	1352756 > 1352730	TGTAAGTACAGATTGTGTAATTCAACG
WGAP106_F	1353905 > 1353931	ACTTCATTAATCTCATCATTAGATGGC
WGAP106_R	1354116 > 1354094	GCAGCAGTTTTATCTGGTAATCG
WTHIN2864_3R	1354685 > 1354664	TGGAATACTTGGAGATGACAGG
WTHIN2864_1F	1361969 > 1361993	CATCAGCATTAAAGTAACCTTGTCC
WTHIN2864_1R	1362667 > 1362646	CACAGCCTCTTAGATGTGTACC
WGAP113_F	1363692 > 1363717	CAATCATCATAGACTCAACTTACCAC
WGAP113_2R	1364243 > 1364219	TGAAATGTATACTCTAGACTGGGAC
WGAP113_R	1364783 > 1364755	GCTATTAAGTGGTATATGTGTTTATCAGG
WGAP54_F	1369934 > 1369962	TGACAATATATCACCTGACTATTAACCTC
WGAP54_R	1370929 > 1370906	TAGAGTTTGTATTGGGAGATCGTG
WTHIN2864_5R	1372623 > 1372598	GAAGTCATTAATTAGTCAATGGAGGG
WTHIN2864_6F	1376372 > 1376392	ACCTACCACAAACGCTATACC
WTHIN2864_6R	1377020 > 1376993	TGTTACTCAATAGTTAAGAATGATGTGC
WTHIN2864_7F	1381544 > 1381572	AATGTAGCAATAACAATAAGAAAGAGG
WTHIN2864_7R	1382346 > 1382321	GATGTAATGAAGGAGAATTCTAACGG
WTHIN_1383570_F	1383645 > 1383665	AGTACCATTTACTCCACCTGC
WGAP9_F	1384020 > 1384049	CCAAATATTGTATAAACTCTACACTTTCTC
WGAP9_R	1385340 > 1385314	TTTGATAGTGTGCTCGATATTATACC
WTHIN2864_9F	1392168 > 1392190	AATACCCTTCTTTCACAACAAGC
WGAP8_F	1392370 > 1392396	TTAACACATACCCAATAATACTGAAGG
WGAP8_R	1392900 > 1392872	CAGTTGAAAATACTCTGTTAATTACTGG



Primer name	Position	Primer sequence (5' > 3')
WGAP7_F	1395735 > 1395763	ATTACGAAGCTATTACATTAGATGAAGTC
WGAP7_R	1396093 > 1396064	ATAGGGTGTATATATCTGTATAGATGGTAC
758_RC2_F	1396151 > 1396180	GATTATGTCAAAAAC TAGTCATCTTTTAGC
758_RC2_R	1397602 > 1397578	GGATATAGTAAGTCAGTGTACTGGC
WGAP6_F	1402487 > 1402509	CTACTGGTATCTCCATACCAACC
WGAP6_R	1402652 > 1402634	CAATTCCAGCACCATTGGG
WTHIN2864_10F	1403051 > 1403071	ACAACAAAGCTGTTTTACACC
WTHIN2864_10R	1404039 > 140401	CTGAATCAGAAGATGAAGAATTGTGC
WTHIN2864_11F	1414133 > 1414157	AGTGATATAGCAATTGATGGAATGG
WGAP5_F	1414867 > 1414894	ACATGATTGAGAATACCTTACAAACAAC
WGAP5_R	1415671 > 1415649	ATTGTTTGTGATGGGAATATGGG
WTHIN_1427957_F	1428033 > 1428056	AGACACGAATCAGGTAAATATCCC
WGAP4_F	1428287 > 1428311	GATGAAACCACAGAACAAAAAGTCC
WGAP4_R	1428556 > 1428537	AGATGCTCTCTGCATGTTGG
WGAP3_F	1438066 > 1438092	GAATATGATGCTATAATAGCTGCTCAC
WTHIN2864_14F	1439495 > 1439515	AGACCCAAGATCTACAACAGC
WGAP3_R	1439898 > 1439867	CATGGTAATAATTAAGATAATTTCTGGTGTG
WTHIN2864_14R	1440686 > 1440658	GATGTTAATGGGAAGATTAGAGATTATGC
WTHIN2864_15F	1447197 > 1447221	AGAGACAGAAGACTCAGTATTTACC
WTHIN2864_15R	1448011 > 1447989	TCGCTGTGATAAAAAGCTCTAAGG
WL2TP1_check_2F	1449621 > 1449649	TCTCAACTATAAACAATAGAGAATTTGGC
WL2TP1_check_2R	1450313 > 1450291	GCTAGACAAGTTTTGATGTTGGC
WL2TP1_check_1F	1450778 > 1450805	TCACCCATAAATTCTATCATAATTACAGC
WTHIN2864_16F	1450873 > 1450898	AGAACACTTGTTTACATATTGCTTGC
WTHIN2864_16R	1452277 > 1452252	GCAGGCATTAGATAATCTACAAGAGG
WGAP2_F	1458669 > 1458689	ATTTCTCCAGTCGTAGTCTC
WGAP2_R	1458869 > 1458850	GCTGCATTAACACCTAGAGC
WGAP1_F	1463837 > 1463860	CTTAACACTTACACCAATGCCAC
WGAP1_R	1464291 > 1464271	GGCCAGATGAATCTTCGACTG
WTHIN2864_18F	1466403 > 1466423	ACAACAAGCAAATGTAGCACC
WTHIN2864_18R	1466907 > 1466882	AGTGTTCATGAAACTTTGTATGG
WTHIN2864_19F	1471341 > 1471365	CTGAACCAATAGTAAAACCTTGCAGC
WTHIN2864_19R	1472505 > 1472477	GAAATAGCATCATAAGAAAGTACTAGGTG
WTHIN2864_21F	1487529 > 1487552	CCCATTAGTAACATCTGCAATAGC
WTHIN2864_21R	1488337 > 1488311	GTCATGAACATATTAATGTGTGTCTGC
WTHIN2864_22F	1489822 > 1489842	GGTGGTGAACCTGGAGTAAGG
WTHIN2864_22R	1490915 > 1490890	TCAGAATTGATTGAGTAACTTATGCC
WTHIN2864_23F	1491063 > 1491087	CTTTGGTATAGAAATTGGAGGAAGG
WTHIN2864_23R	1492041 > 1492016	GTGTTTCTGTTTAAAGGATAAGATGG
WTHIN2864_24F	1492872 > 1492895	AGCATAGTACGAATACTTAGTGGG
WTHIN2864_24R	1493373 > 1493349	CTGGTTTAACTAAGGGTGTATGG
WGAP55_F	1498494 > 1498522	ACTGATAGAACATATAACAACATCACCG
WGAP55walk_1F	1499373 > 1499395	AGTACACTGCTTTCTCATACCAC
WGAP55walk_1R	1500174 > 1500152	GTGAAATGTATGCGTAGTAGTTC
WGAP55_R	1500710 > 1500686	TGGACCTATAGAATTGGTTTACTGC
WGAP56_F	1505871 > 1505894	CAGAAAATGCAGAAGAAATTGCAC
WGAP56_R	1506892 > 1506868	GTATTTCTATAATGTGCGCATGACCC
WTHIN2864_26F	1507985 > 1508005	GTTTGGTAAAGTTCGCAGAGC
WTHIN2864_26R	1509300 > 1509276	CGATTTCTATTACAGGATTTGCAGG
WGAP57_F	1512265 > 1512287	GAAAATCTACCAAACGAGAGGG
WGAP57_R	1513153 > 1513122	ATGCTTATTATAGTGATAATGTTGGATATCAG
WGAP58_F	1514697 > 1514716	ATAGATGCCCCCATCAAAGC
WGAP58_R	1515427 > 1515404	AACATGTGGTGCAATTTATACAGG

Duplicates	
WGAP48R (= 22799_R)	24993_557_F (= WTHIN440_5F)
WGAP35F (= 16700_F)	24993_557_R (= WTHIN440_5R)
17139_F (= WGAP34F)	WL2TP1_check_1R (= WTHIN2864_16R)
3687_RI (= WGAP49_2F)	WTHIN1674_16R (= WTHIN1174_1R)

C2: Primers designed to amplify the *tuf* and *rho* regions.

Primer name	Primer sequence (5' > 3')
tuf_1	GCAAACAGGTGGTGCTGG
tuf_2	CATTTTCTTGCGCATAGACTCC
tuf_3	CCAGGATCTTGACACTGACC
tuf_4	TCCATAACACCAATATCCTGC
rho_1	ACACCTGTTGCACGTCG
rho_2	ACAAAGCAAGCCATGAAGC
rho_3	GACAACCTGAACATGCTCC
rho_4	CCTATCCATTCTCCAATCTTTTGC

C3: Primers designed to amplify and clone ORFs into the pCMViUBs and TOPO[®] pET vectors.

Restriction enzyme sites are underlined.

Primer name	Primer sequence (5' > 3')
3630F_BamHI	<u>GGATCCT</u> TTTACATTACAAAAACAATTTAACAGTAC
3630R_Sall	<u>GTCGACT</u> TTACTGTCATGCCCT
4470F_BamHI	<u>GGATCCA</u> ATGATTCTCATTGTCTGGT
4470R_Sall	<u>GTCGACT</u> TAAAACCTTAAACTTTGTACCTATCAA
5270F_BamHI	<u>GGATCCA</u> TGTTTACTTTGCCAGAAGTGC
5270R_Sall	<u>GTCGACT</u> TATTTAACATTATCAATACATTGAGAA
5400F_BamHI	<u>GGATCCA</u> TGCAAAACGTAATAATATATTGTTTTG
5400R_Sall	<u>GTCGACT</u> CAATAAGTATTTAATACTAATGTATTACC
5430F_BamHI	<u>GGATCCA</u> TGTTTGAATCTTTAACTAGTAGTTTAAAC
5430R_Sall	<u>GTCGACT</u> TATTCATTGTTTTTCAGTAAATTCAT
7300F_BamHI	<u>GGATCCA</u> TGAATCAGCAAATGGTAGTG
7300R_Sall	<u>GTCGACT</u> CACTCATGATTAACACCAC
8050F_BamHI	<u>GGATCCT</u> CTGAAGATATTGAGCAATATGATC
8050R_Sall	<u>GTCGACT</u> TACTTCTTTAACTTAAACAGGAATAAATATTG
pET3630F	CACCTTTTTTACATTACAAAAACAATTTAACAGTACAAC
pET3630R	CACTGCATGCCCTATGTAAC
pET4470F	CACCTAATGATTCTCATTGTCTGGTAAT
pET4470R	AAACTTAAACTTTGTACCTATCAA
pET5270F	CACCATGTTTACTTTGCCAGAAGTGC
pET5270R	TTTAACATTATCAATACATTGAGAAAATC
pET5400F	CACCATGCAAAACGTAATAATATATTGTTTTTGG
pET5400R	ATAAGTATTTAATACTAATGTATTACCACTC
pET5430F	CACCATGTTTGAATCTTTAACTAGTAGTTTAAAC
pET5430R	TTCATTGTTTTTCAGTAAATTCATAAAAT
pET7300F	CACCATGCCTGAGCAAATGTATC
pET7300R	TAACATCATGATTAACACCACGTCG

C4: Vector specific primers used in this study.

Primer name	Primer sequence (5' > 3')
pET TrxFus Forward	TTCCTCGACGCTAACCTG
pET T7 Reverse	TAGTTATTGCTCAGCGGTGG
pCMViUBs IECO	GGCTAGCCTCGAGAATTC
pCMViUBs CMV991	CAGGGATGCCACCCGGG
pGEM SP6	ATTTAGGTGACACTATAG
pGEM T7	TAATACGACTCACTATAGGG

Appendix D: Protein classification scheme

- 0.0.0 Unknown function, no known homologs
- 0.0.1 Conserved in Rickettsiales
- 0.0.2 Conserved in organism other than Rickettsiales

- 1.0.0 Cell processes
 - 1.1.1 Chemotaxis and mobility
 - 1.2.1 Chromosome replication
 - 1.3.1 Chaperones
 - 1.4.0 Protection responses
 - 1.4.1 Cell killing
 - 1.4.2 Detoxification
 - 1.4.3 Drug/analog sensitivity
 - 1.4.4 Radiation sensitivity
 - 1.5.0 Transport/binding proteins
 - 1.5.1 Amino acids and amines
 - 1.5.2 Cations
 - 1.5.3 Carbohydrates, organic acids, alcohols
 - 1.5.4 Anions
 - 1.5.5 Other
 - 1.5.6 Type IV secretion
 - 1.5.7 ABC transporters
 - 1.6.0 Adaptation
 - 1.6.1 Adaptations, atypical conditions
 - 1.6.2 Osmotic adaptation
 - 1.6.3 Fe storage
 - 1.7.1 Cell division

- 2.0.0 Macromolecule metabolism
 - 2.1.0 Macromolecule degradation
 - 2.1.1 Degradation of DNA
 - 2.1.2 Degradation of RNA
 - 2.1.3 Degradation of polysaccharides
 - 2.1.4 Degradation of proteins, peptides, glycoproteins
 - 2.2.0 Macromolecule synthesis, modification
 - 2.2.01 Amino acyl tRNA synthesis, tRNA modification
 - 2.2.02 Basic proteins - synthesis, modification
 - 2.2.03 DNA - replication, repair, modification
 - 2.2.04 Glycoprotein
 - 2.2.05 Lipopolysaccharide
 - 2.2.06 Lipoprotein
 - 2.2.07 Phospholipids
 - 2.2.08 Polysaccharides - (cytoplasmic)
 - 2.2.09 Protein modification
 - 2.2.10 Proteins, translation and modification
 - 2.2.11 RNA synthesis, modification, DNA transcription
 - 2.2.12 tRNA

- 3.0.0 Metabolism of small molecules
 - 3.1.0 Amino acid biosynthesis
 - 3.1.01 Alanine
 - 3.1.02 Arginine
 - 3.1.03 Asparagine
 - 3.1.04 Aspartate
 - 3.1.05 Chorismate
 - 3.1.06 Cysteine
 - 3.1.07 Glutamate
 - 3.1.08 Glutamine
 - 3.1.09 Glycine
 - 3.1.10 Histidine
 - 3.1.11 Isoleucine
 - 3.1.12 Leucine
 - 3.1.13 Lysine
 - 3.1.14 Methionine
 - 3.1.15 Phenylalanine
 - 3.1.16 Proline
 - 3.1.17 Serine
 - 3.1.18 Threonine
 - 3.1.19 Tryptophan
 - 3.1.20 Tyrosine
 - 3.1.21 Valine
 - 3.2.0 Biosynthesis of cofactors, carriers
 - 3.2.01 Acyl carrier protein (ACP)
 - 3.2.02 Biotin
 - 3.2.03 Cobalamin
 - 3.2.04 Enterochelin
 - 3.2.05 Folic acid
 - 3.2.06 Heme, porphyrin
 - 3.2.07 Lipoate
 - 3.2.08 Menaquinone, ubiquinone
 - 3.2.09 Molybdopterin
 - 3.2.10 Pantothenate
 - 3.2.11 Pyridine nucleotide
 - 3.2.12 Pyridoxine
 - 3.2.13 Riboflavin
 - 3.2.14 Thiamin
 - 3.2.15 Thioredoxin, glutaredoxin, glutathione
 - 3.2.16 Biotin carboxyl carrier protein (BCCP)



- 3.3.0 Central intermediary metabolism
 - 3.3.01 2'-Deoxyribonucleotide metabolism
 - 3.3.02 Amino sugars
 - 3.3.03 Entner-Doudoroff
 - 3.3.04 Gluconeogenesis
 - 3.3.05 Glyoxylate bypass
 - 3.3.06 Incorporation metal ions
 - 3.3.07 Miscellaneous glucose metabolism
 - 3.3.08 Miscellaneous glycerol metabolism
 - 3.3.09 Non-oxidative branch, pentose pathway
 - 3.3.10 Nucleotide hydrolysis
 - 3.3.11 Nucleotide interconversions
 - 3.3.12 Oligosaccharides
 - 3.3.13 Phosphorus compounds
 - 3.3.14 Polyamine biosynthesis
 - 3.3.15 Pool, multipurpose conversions of intermediary metabolism
 - 3.3.16 S-adenosyl methionine
 - 3.3.17 Salvage of nucleosides and nucleotides
 - 3.3.18 Sugar-nucleotide biosynthesis, conversions
 - 3.3.19 Sulfur metabolism
 - 3.3.20 amino acids
 - 3.3.00 other
- 3.4.0 Degradation of small molecules
 - 3.4.1 Amines
 - 3.4.2 Amino acids
 - 3.4.3 Carbon compounds
 - 3.4.4 Fatty acids
 - 3.4.5 Other
- 3.5.0 Energy metabolism, carbon
 - 3.5.1 Aerobic respiration
 - 3.5.2 Anaerobic respiration
 - 3.5.3 Electron transport
 - 3.5.4 Fermentation
 - 3.5.5 Glycolysis
 - 3.5.6 Oxidative branch, pentose pathway
 - 3.5.7 Pyruvate dehydrogenase
 - 3.5.8 TCA cycle
- 3.6.0 Fatty acid biosynthesis
 - 3.6.1 Fatty acid and phosphatidic acid biosynthesis
- 3.7.0 Nucleotide biosynthesis
 - 3.7.1 Purine ribonucleotide biosynthesis
 - 3.7.2 Pyrimidine ribonucleotide biosynthesis
- 4.0.0 Cell envelope
 - 4.1.0 Periplasmic/exported/lipoproteins
 - 4.1.1 Inner membrane
 - 4.1.2 Murein sacculus, peptidoglycan
 - 4.1.3 Outer membrane constituents
 - 4.1.4 Surface polysaccharides & antigens
 - 4.1.5 Surface structures
 - 4.2.0 Ribosome constituents
 - 4.2.1 Ribosomal and stable RNAs
 - 4.2.2 Ribosomal proteins - synthesis, modification
 - 4.2.3 Ribosomes - maturation and modification
- 5.1.0 Laterally acquired elements
 - 5.1.1 Colicin-related functions
 - 5.1.2 Phage-related functions and prophages
 - 5.1.3 Plasmid-related functions
 - 5.1.4 Transposon/insertion element-related functions
- 6.0.0 Regulation
 - 6.1.1 Global regulatory functions
- 7.0.0 Not classified (included putative assignments)
 - 7.1.1 DNA sites, no gene product
 - 7.2.1 Cryptic genes

Appendix E: *E. ruminantium* gene list

The first column indicates the systematic identification number of each predicted ORF, followed by the gene name, protein product and length in amino acids. Columns 5 to 7 show the transmembrane helices and signal sequences predicted by TMHMM2.0, SignalP3.0 and Phobius (th = transmembrane helix). Columns 8 and 9 represent the subcellular localisation predictions by CELLO and pSORTb2.0: C = cytoplasmic, P = periplasmic, IM = inner membrane, OM = outer membrane, E = extra cellular, U = unknown. In column 10 helix-turn-helix motifs are represented by plus signs. The size and frequency of tandem repeats, the EC number and functional class are given in the last three columns.

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
0010	<i>gapB</i>	NAD(P)-dependent glyceraldehyde 3-phosphate dehydrogenase [†]	335				C	C			1.2.1.59	3.3.15
0020	<i>elbB</i>	enhancing lycopene biosynthesis protein 2*	220				C	C				7.0.0
0030	<i>proC</i>	pyrroline-5-carboxylate reductase	271				OM	U			1.5.1.2	3.1.16
0040	<i>dnaZ</i>	DNA polymerase III, gamma subunit*	487				C	U			2.7.7.7	2.2.03
0050		unknown	101				C	U				7.0.0
0060	<i>asd</i>	aspartate-semialdehyde dehydrogenase	337				C	U			1.2.1.11	3.1.0
0070	<i>metK</i>	S-adenosylmethionine synthetase	401				C	C			2.5.1.6	3.3.16
0080	<i>ubiF</i>	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase*	390	1			C	U		158 bp x 2.8 (C-terminus)	1.14.13.-	3.2.08
0090		membrane protein*	193	1	signal	2 th	P	U				4.1.1
0110	<i>glyQ</i>	glycyl-tRNA synthetase alpha chain	280				C	C			6.1.1.14	2.2.01
0120	<i>glyS</i>	glycyl-tRNA synthetase beta chain	702				C	U			6.1.1.14	2.2.01
0130	<i>dnaJ</i>	chaperone protein DnaJ	382				OM	C				1.3.1
0140	<i>nadC</i>	nicotinate-nucleotide pyrophosphorylase [carboxylating]	277				C	C			2.4.2.19	3.2.11
0150		integral membrane protein*	195	5		5 th	IM	IM				4.1.1
0160	<i>ruvC</i>	crossover junction endodeoxyribonuclease RuvC	160				C	U			3.1.22.4	2.2.03
0170	<i>coxC</i>	cytochrome c oxidase subunit III	274	7		7 th	IM	IM			1.9.3.1	3.5.3
0180	<i>hemE</i>	uroporphyrinogen decarboxylase	335				C	U			4.1.1.37	3.2.06
0190	<i>corC</i>	magnesium and cobalt efflux protein [†]	288			1 th	C	C				1.5.2
0200		protease [†]	178				C	U				5.1.2
0210		genetic exchange protein [†]	394				E	U				5.1.2
0220	<i>bioC</i>	biotin synthesis protein BioC [†]	249				OM	IM				3.2.02
0230	<i>nadA</i>	quinolinate synthetase A	314				C	C			1.4.3.-	3.2.11
0240	<i>fdxA</i>	ferredoxin	125				C	C				3.5.3
0250		unknown	457				C	C		297 bp x 2.8		0.0.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
0260	<i>virD4</i>	type IV secretion system protein VirD4	801	3		2 th	C	IM		6 bp x 5.0		1.5.6
0270	<i>virB11</i>	type IV secretion system protein VirB11	332				C	U				1.5.6
0280	<i>virB10</i>	type IV secretion system protein VirB10	448	1		1 th	P	U		6 bp x 9.0		1.5.6
0290	<i>virB9</i>	type IV secretion system protein VirB9	267		signal	signal	E	U				1.5.6
0300	<i>virB8</i>	type IV secretion system protein VirB8	232	1		1 th	OM	U				1.5.6
0310		riboflavin biosynthesis protein*	371				C	C			3.5.4.25	3.2.13
0320		unknown	354				C	U				7.0.0
0330		integral membrane protein*	159	2 ⁺		2 th	C	C				4.1.1
0340	<i>dapF</i>	diaminopimelate epimerase	265				C	C			5.1.1.7	3.1.13
0350		Unknown	143				C	U				0.0.0
0360	<i>pgk</i>	phosphoglycerate kinase	395				C	C			2.7.2.3	3.3.15
0370	<i>xseA</i>	exodeoxyribonuclease VII large subunit	388				C	U		203 bp x 3.0 (N-terminus)	3.1.11.6	2.1.1
0380		membrane protein*	222	1		1 th	C	U		203 bp x 3.0 (C-terminus)		4.1.1
0390	<i>dapD</i>	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	284				C	IM			2.3.1.117	3.1.13
0400	<i>trmE</i>	tRNA modification GTPase*	439				C	C				2.2.12
0410	<i>dfp</i>	DNA/pantothenate metabolism flavoprotein*	181			signal	C	U			4.1.1.36, 6.3.2.5	2.2.03
0420	<i>recG</i>	ATP-dependent DNA helicase RecG	677			2 th	C	IM			3.6.1.-	2.2.03
0430		NADH-ubiquinone oxidoreductase subunit [†]	320	2		1 th	C	IM		283 bp x 3.2 (C-terminus)	1.6.99.3	3.5.3
0440	<i>dksA</i>	DnaK suppressor protein*	151				C	C				1.6.1
0450	<i>ccmB</i>	heme exporter protein B [†]	220	5		6 th	IM	IM				1.5.7
0460		cation efflux system protein*	306	6		6 th	IM	IM				1.5.2
0470		exported protein*	208		signal	signal	C	C				4.1.0
0480	<i>rpsT</i>	30S ribosomal protein S20	95			signal	C	U				4.2.2
0490	<i>polA</i>	DNA polymerase I	865				OM	C	+	198 bp x 3.0 (C-terminus)	2.7.7.7	2.2.03
0500		unknown	102				C	U				0.0.0
0510	<i>argF</i>	ornithine carbamoyltransferase	305				C	C			2.1.3.3	3.1.02
0520	<i>recF</i>	DNA replication and repair protein RecF*	372				OM	C				2.2.03
0530		uracil DNA glycosylase [†]	263			2 th	C	U			3.2.2.-	2.2.03
0540	<i>def1</i>	peptide deformylase 1*	181				C	C			3.5.1.88	2.2.09
0550	<i>plsC</i>	1-acyl-sn-glycerol-3-phosphate acyltransferase*	241	3		2 th	IM	IM			2.3.1.51	3.6.1
0560	<i>rpe</i>	ribulose-phosphate 3-epimerase	215				C	U			5.1.3.1	3.3.09
0570		integral membrane protein*	265	6		7 th	IM	IM				4.1.1
0580		ABC transporter, ATP binding protein*	239				C	U				1.5.7
0590		integral membrane protein*	613	3		3 th	OM	U				4.1.1
0600	<i>ispB</i>	octaprenyl-diphosphate synthase	325				C	C			2.5.1.-	3.2.08

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
0610	<i>glnA</i>	glutamine synthetase	470				OM	C			6.3.1.2	3.1.08
0620	<i>tyrS</i>	tyrosyl-tRNA synthetase	418				C	C			6.1.1.1	2.2.01
0630	<i>hemA</i>	5-aminolevulinic acid synthase	398				C	C			2.3.1.37	3.2.06
0631		conserved hypothetical protein	122			2 th	C	IM				0.0.2
0640	<i>secF</i>	protein-export membrane protein SecF	289	6		6 th	IM	IM				1.5.5
0650	<i>fbaB</i>	fructose-bisphosphate aldolase class I*	300				C	C			4.1.2.13	3.3.04
0660		unknown	3715				OM	U		300 bp x 2.6, 171 bp x 2.4, 471 bp x 2.7, 171 bp x 2.4		0.0.0
0670	<i>pdhC</i>	dihydrolipoamide acetyltransferase, E2 component of pyruvate dehydrogenase complex	406				C	C			2.3.1.12	3.5.7
0680		unknown	540				OM	U				0.0.0
0690		unknown	470			1 th	OM	U				0.0.0
0700		integral membrane protein*	547	2		2 th, signal	OM	U				4.1.1
0710		unknown	123			signal	C	U				0.0.0
0720		unknown	931			1 th, signal	OM	U				0.0.0
0730		unknown	93				C	U				0.0.0
0740	<i>guaA</i>	GMP synthase [glutamine-hydrolyzing]	528				C	U		170 bp x 4.0 (C-terminus)	6.3.5.2	3.7.1
0750	<i>gltA</i>	citrate synthase	415				C	C			2.3.3.1	3.5.8
0770	<i>gshA</i>	gamma-glutamylcysteine synthetase [†]	399				OM	U				3.2.18
0780	<i>valS</i>	valyl-tRNA synthetase	810				C	C		336 bp x 2.9 (C-terminus)	6.1.1.9	2.2.01
0790	<i>smpB</i>	SsrA-binding protein	148				C	U				2.2.10
0800	<i>ribB</i>	3,4-dihydroxy-2-butanone 4-phosphate synthase	211				C	C				3.2.13
0810	<i>greA</i>	transcription elongation factor GreA	162				C	U				2.2.11
0820	<i>atpA</i>	ATP synthase alpha chain	507				OM	C			3.6.3.14	3.5.9
0830	<i>atpH</i>	ATP synthase delta chain*	189				OM	U			3.6.3.14	3.5.9
0831		integral membrane protein*	84	2		2 th	IM	U				4.1.1
0840		integral membrane protein*	413	2		1 th, signal	C	C				4.1.1
0850		membrane protein*	258	1		1 th	C	C	+			4.1.1
0860	<i>lolE</i>	lipoprotein releasing system transmembrane protein LolE*	411	4		4 th	IM	IM	+			1.5.7
0870		conserved hypothetical protein	339				C	U				0.0.1
0880	<i>ccmF</i>	cytochrome c-type biogenesis protein CcmF	638	14		15 th	IM	IM				2.2.13
0890		aminomethyl transferase*	280				OM	U			2.1.2.10	3.3.00
0900	<i>purF</i>	glutamine phosphoribosylpyrophosphate amidotransferase	466				C	U			2.4.2.14	3.7.1
0910	<i>pth</i>	peptidyl-tRNA hydrolase	193				C	U			3.1.1.29	2.2.01
0920	<i>rplY</i>	50S ribosomal protein L25*	208				C	U				4.2.2
0930	<i>comF</i>	competence protein F [†]	230				C	U				7.0.0
0940	<i>dapE</i>	succinyl-diaminopimelate desuccinylase*	383				E	C			3.5.1.18	3.1.13

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
0950		glutathione-regulated potassium-efflux system protein*	569	10		13 th	IM	IM				1.5.2
0960		conserved hypothetical protein	193				P	U				0.0.2
0970		integral membrane protein*	155	4		4 th	IM	IM				4.1.1
0980	<i>pdhB</i>	pyruvate dehydrogenase E1 component, beta subunit*	332				C	C			1.2.4.1	3.5.7
0990		integral membrane protein*	607	2		2 th	OM	OM				4.1.1
1000	<i>tldD</i>	TldD protein	475				OM	C				6.0.0
1010		conserved hypothetical GTP-binding protein	363				C	C				0.0.2
1020	<i>ispF</i>	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	173				C	U		154 bp x 5.1 (C-terminus)	4.6.1.12	3.2.08
1030	<i>ispD</i>	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase*	242			signal	C	U			2.7.7.60	3.2.08
1040		integral membrane protein*	1165	2		2 th	E	C		294 bp x 2.8		4.1.1
1050		integral membrane protein*	454	2		2 th	OM	U				4.1.1
1060	<i>purE</i>	phosphoribosylaminoimidazole carboxylase catalytic subunit	170				E	U			4.1.1.21	3.7.1
1070		exported protein*	180		signal	1 th	C	U				4.1.0
1080	<i>ihfA</i>	integration host factor alpha-subunit*	99				C	C				2.2.03
1090		conserved hypothetical protein	122				C	U	+			0.0.2
1100		unknown	161				E	U				0.0.0
1110		unknown	661				E	C		2 bp x 2.5, 27 bp x 56.0		0.0.0
1120	<i>trpS</i>	tryptophanyl-tRNA synthetase	332				OM	C			6.1.1.2	2.2.01
1130	<i>grpE</i>	GrpE protein	199				C	C				1.3.1
1140	<i>ribD</i>	riboflavin biosynthesis protein RibD	365				C	C			3.5.4.26, 1.1.1.193	3.2.13
1150		unknown	179				CP	C				0.0.0
1160	<i>pyrG</i>	CTP synthase	540			signal	C	U			6.3.4.2	3.7.2
1170	<i>secG</i>	protein-export membrane protein SecG*	110	2		1 th, signal	P	U				1.5.0
1180		integrase/recombinase XerD or XerC*	312				C	C				2.2.03
1190	<i>lolD</i>	lipoprotein releasing system ATP-binding protein LolD	228			signal	C	IM				1.5.7
1200	<i>maeB</i>	NADP-dependent malic enzyme	755				IM	IM			1.1.1.40	3.3.15
1210		exported protein*	877		signal	signal	OM	OM				4.1.0
1220	<i>lnt</i>	apolipoprotein N-acyltransferase*	506	7 [‡]		7 th	IM	IM			2.3.1.-	2.2.06
1230		unknown	186				C	IM		237 bp x 2.4		0.0.0
1240		NADH-quinone oxidoreductase subunit*	492	13		14 th	IM	IM			1.6.99.5	3.5.3
1250		membrane protein*	99	1		1 th	P	U				4.1.1
1260		membrane protein*	149	1 [‡]		1 th	C	U				4.1.1
1270		unknown	93				P	U				7.0.0
1280		conserved hypothetical protein	153				C	U				0.0.2
1290		unknown	564				E	U				0.0.0
1300		unknown	1334				E	OM				7.0.0
1310	<i>fbpA</i>	iron-binding periplasmic protein*	348		signal	signal	C	P				1.5.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
1320	<i>rpsP</i>	30S ribosomal protein S16	87				C	U				4.2.2
1330	<i>proP</i>	proline/betaine transporter	422	12		12 th	IM	IM				1.5.1
1340		conserved hypothetical protein	55				C	U				0.0.2
1350		short chain dehydrogenase*	231				C	U			1.1.1.-	7.0.0
1360	<i>pheS</i>	phenylalanyl-tRNA synthetase alpha chain	344				C	C			6.1.1.20	2.2.01
1370	<i>rplT</i>	50S ribosomal subunit protein L20	123				C	U				4.2.2
1380	<i>rplM</i>	50S ribosomal protein L35	66				C	U				4.2.2
1390		conserved hypothetical protein	221				C	U				0.0.1
1400	<i>rho1</i>	transcription termination factor 1	478				C	C				2.2.11
1410		unknown	80				C	U				7.0.0
1420		dihydrolipoamide dehydrogenase, E3 component of pyruvate or 2-oxoglutarate dehydrogenase complex*	469				C	C			1.8.1.4	3.5.0
1430		unknown	951			1 th	OM	U		198 bp x 2.3		0.0.0
1440		membrane protein*	482	1		1 th	C	IM				4.1.1
1450		membrane protein*	208	1		1 th	OM	U				4.1.1
1460		exported protein*	180		signal	signal	C	C				4.1.0
1470		unknown	262				C	U				7.0.0
1480		truncated glutamine synthetase [†]	268				C	C			6.3.1.2	3.1.08
1490		ABC transporter, membrane-spanning protein [†]	374	8		9 th	IM	IM				1.5.7
1500	<i>alaS</i>	alanyl-tRNA synthetase	887				C	C			6.1.1.7	2.2.01
1510	<i>sucD</i>	succinyl-CoA synthetase, alpha subunit	295				C	U			6.2.1.5	3.5.8
1520	<i>sucC</i>	succinyl-CoA synthetase, beta subunit	386				C	U			6.2.1.5	3.5.8
1530	<i>rpsU</i>	30S ribosomal protein S21 [†]	112	1		1 th	C	U				4.2.2
1540		exported protein*	342		signal	signal	C	U				4.1.0
1550	<i>map2</i>	major antigenic protein 2	209	1 [†]		1 th	P	U				4.1.3
1560		2-nitropropane dioxygenase*	345				C	U				3.3.0
1570		cytochrome b561*	173	5 [†]		5 th	IM	IM				3.5.3
1580		ABC transporter, membrane-spanning protein*	536	12		13 th	IM	IM				1.5.7
1590		secretion protein*	514	1		2 th	OM	IM				1.5.5
1600		unknown	204				C	U				0.0.0
1610		conserved hypothetical protein	542				C	U				0.0.2
1620		integral membrane protein*	197	4		4 th	IM	IM				4.1.1
1630	<i>rpsL</i>	30S ribosomal protein S12	123				P	U				4.2.2
1640	<i>rpsG</i>	30S ribosomal protein S7	160				C	U				4.2.2
1650	<i>fusA</i>	elongation factor G	689				C	C				2.2.10
1660	<i>tufA</i>	elongation factor Tu-A	395				C	C				2.2.10
1670	<i>nusG</i>	transcription antitermination protein NusG	179				C	C				2.2.11
1680	<i>rplK</i>	50S ribosomal protein L11	147				P	U				4.2.2
1690	<i>rplA</i>	50S ribosomal protein L1	220				C	U				4.2.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
1700	<i>rpIJ</i>	50S ribosomal protein L10	160				C	IM				4.2.2
1710	<i>rpIL</i>	50S ribosomal protein L7/L12	131				C	C				4.2.2
1720	<i>rpoB</i>	DNA-directed RNA polymerase beta chain	1380				C	C			2.7.7.6	2.2.11
1730	<i>rpoC</i>	DNA-directed RNA polymerase beta' chain	1411				C	C			2.7.7.6	2.2.11
1740	<i>bioF</i>	8-amino-7-oxononanoate synthase*	367				C	IM			2.3.1.47	3.2.02
1750		integral membrane protein*	142	4		4 th	IM	IM				4.1.1
1760	<i>rnhB</i>	ribonuclease HII	212				C	U		208 bp x 3.6 (C-terminus)	3.1.26.4	2.1.2
1770		unknown	1529				OM	OM				0.0.0
1780		Na ⁺ /H ⁺ antiporter subunit [†]	172	2		2 th	IM	IM				1.5.2
1790		membrane protein*	205	1		1 th	C	U				4.1.1
1800		unknown	257				IM	U				7.0.0
1810	<i>pyrD</i>	dihydroorotate dehydrogenase	346				OM	U			1.3.3.1	3.7.2
1820	<i>def2</i>	peptide deformylase 2*	194				C	C			3.5.1.88	2.2.09
1830	<i>argH</i>	argininosuccinate lyase	462				C	U		202 bp x 2.0 (N-terminus)	4.3.2.1	3.1.02
1840		unknown	267				C	U				7.0.0
1850	<i>pdxH</i>	pyridoxamine 5'-phosphatase oxidase	194				C	U			1.4.3.5	3.2.12
1851		unknown	92				C	C				0.0.0
1860		membrane protein*	270	1 [†]		signal	C	U				4.1.1
1870	<i>dnaE</i>	DNA polymerase III, alpha subunit	1119				C	C			2.7.7.7	2.2.03
1880	<i>aroE</i>	3-phosphoshikimate 1-carboxyvinyltransferase	427			1 th	IM	IM			2.5.1.19	3.1.05
1890	<i>sdhC</i>	succinate dehydrogenase cytochrome b-556 subunit*	132	3		3 th	IM	IM			1.3.5.1	3.5.3
1891	<i>sdhD</i>	succinate dehydrogenase cytochrome b small subunit*	116	3		3 th	IM	IM			1.3.5.1	3.5.3
1900		unknown	417				OM	U				0.0.0
1910	<i>thiD</i>	phosphomethylpyrimidine kinase*	266				C	U			2.7.4.7	3.2.14
1920		conserved hypothetical protein	230				C	U				0.0.2
1930		integral membrane protein*	373	8		7 th	IM	IM				4.1.1
1940	<i>rpsD</i>	30S ribosomal protein S4	202				P	U				4.2.2
1950		conserved hypothetical protein	69				C	C				0.0.2
1960		exported protein*	383		signal	signal	OM	U				4.1.0
1970		acetyltransferase [†]	262				C	U			2.3.1.-	7.0.0
1980	<i>pgpA</i>	phosphatidylglycerophosphatase A*	168	4		4 th	IM	IM			3.1.3.27	2.2.07
1990	<i>tig</i>	trigger factor	446				C	U				1.3.1
2000	<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	198				C	U			3.4.21.92	2.1.4
2010	<i>clpX</i>	ATP-dependent Clp protease ATP-binding subunit ClpX	406				C	C				2.1.4
2020	<i>lon</i>	ATP-dependent protease La	801			signal	C	C			3.4.21.53	2.1.4
2030	<i>fmt</i>	methionyl-tRNA formyltransferase	303				C	U			2.1.2.9	2.2.01
2040		conserved hypothetical protein	272				C	U				0.0.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
2050		conserved hypothetical protein	285				C	C				0.0.2
2060	<i>thiE</i>	thiamine-phosphate pyrophosphorylase*	350				C	U			2.5.1.3	3.2.14
2070		integral membrane protein*	431	4 [†]		2 th, signal	IM	IM				4.1.1
2080		integral membrane protein*	94	2		3 th	IM	U				4.1.1
2090	<i>ftsK</i>	cell division protein FtsK*	855	5 [†]		4 th	IM	IM		45 bp x 4.1		1.7.1
2100		integral membrane protein*	307	6		6 th	IM	IM				4.1.1
2110	<i>argD</i>	acetylornithine/succinyl-diaminopimelate aminotransferase	391				C	U			2.6.1.11, 2.6.1.17	3.1.0
2120		histidine kinase sensor component of a two-component regulatory system [†]	477	2 [†]		1 th, signal	OM	IM				6.1.2
2130	<i>mutL</i>	DNA mismatch repair protein MutL	689				C	U				2.2.03
2140	<i>smf</i>	DNA processing protein chain A*	375				C	U				2.2.03
2150	<i>fabF</i>	3-oxoacyl-[acyl-carrier-protein] synthase II	423				C	IM			2.3.1.41	3.6.1
2160	<i>acpP</i>	acyl carrier protein	92				C	C				3.2.01
2170		unknown	1073				E	OM		252 bp x 2.7		0.0.0
2180		integral membrane protein*	876	2		2 th	OM	C				4.1.1
2190	<i>rpmG</i>	50S ribosomal protein L33	56			1 th	C	U				4.2.2
2200		integral membrane protein*	235	6		6 th	IM	IM				4.1.1
2210	<i>dsbB</i>	disulfide bond formation protein B [†]	160	4		4 th	IM	IM				2.2.09
2220		unknown	170			1 th	C	U				7.0.0
2230	<i>trmU</i>	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	370				E	U			2.1.1.61	2.2.01
2240		membrane protein*	369	1 [†]		1 th	C	U				4.1.1
2250		membrane protein*	347	1 [†]		1 th	IM	U				4.1.1
2260		membrane protein*	313	1 [†]		1 th	E	U				4.1.1
2270		membrane protein*	384	1 [†]		1 th	E	U				4.1.1
2280		membrane protein*	341	1 [†]		1 th	C	U				4.1.1
2290		membrane protein*	342	1 [†]		1 th	E	U				4.1.1
2300		membrane protein*	370	1		1 th	C	U	+			4.1.1
2310		exported protein*	317		signal	1 th	C	C				4.1.0
2320		exported protein*	307		signal	1 th	C	U				4.1.0
2330		membrane protein*	306	1 [†]		1 th	IM	U				4.1.1
2340		membrane protein*	326	1 [†]		1 th	C	U				4.1.1
2370		unknown	417				E	U				0.0.0
2380		unknown	332			signal	OM	U				0.0.0
2390	<i>uvrD</i>	DNA helicase II	639				C	U			3.6.1.-	2.2.03
2400		membrane protein*	391	1 [†]		1 th	E	U		90 bp x 2.0		4.1.1
2410		membrane protein*	326	1 [†]		1 th	C	U				4.1.1
2420	<i>gyrA</i>	DNA gyrase subunit A	898				OM	C			5.99.1.3	2.2.03
2430	<i>nth</i>	endonuclease III	210				C	U			4.2.99.18	2.2.03

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
2440		integral membrane protein*	220	2		2 th	IM	U				4.1.1
2450	<i>htpG</i>	chaperone protein HtpG	637				C	C				1.3.1
2460	<i>purB</i>	adenylosuccinate lyase	432				C	C			4.3.2.2	3.7.1
2470		integral membrane protein*	358	2		4 th	C	IM				4.1.1
2480		integral membrane protein*	368	3		4 th	IM	IM				4.1.1
2490		unknown	831			3 th	OM	IM				0.0.0
2500		unknown	305				OM	U				0.0.0
2510		unknown	807			2 th	OM	OM				0.0.0
2520		biotin-[acetyl-CoA-carboxylase] synthetase*	252				E	U			6.3.4.15	3.2.02
2530		glutathione S-transferase*	241				C	C		155 bp x 3.0 (C-terminus)	2.5.1.18	3.4.5
2540		exported protein*	175		signal	signal	C	U				4.1.0
2550		ABC transporter, ATP-binding protein*	340				C	C				1.5.7
2560	<i>tatA</i>	Sec-independent protein translocase membrane protein [†]	56	1		1 th	C	U				1.5.5
2570	<i>recR</i>	recombination protein RecR*	195				C	U				2.2.03
2580		ABC transporter, periplasmic solute binding protein*	287	1 [†]		signal	C	P				1.5.7
2590		ABC transporter, ATP-binding protein*	242				C	IM				1.5.7
2600	<i>ubiB</i>	ubiquinone biosynthesis protein UbiB*	480	1		2 th	IM	C		221 bp x 2.0 (C-terminus)		3.2.08
2610		integral membrane protein*	401	12		12 th	IM	IM		221 bp x 2.0 (N-terminus)		4.1.1
2620		conserved hypothetical protein	445				C	C				0.0.2
2630		unknown	1202				OM	OM		375 bp x 2.1 (C-terminus)		0.0.0
2640		conserved hypothetical protein	274				E	U				0.0.1
2650	<i>sucA</i>	2-oxoglutarate dehydrogenase E1 component	913				OM	C			1.2.4.2	3.5.8
2660		unknown	411				E	U				5.1.2
2670	<i>dapA</i>	dihydrodipicolinate synthase	296				OM	U			4.2.1.52	3.1.13
2680		HIT-like protein*	113				C	C				7.0.0
2690		unknown	352				E	U				7.0.0
2700	<i>mutS</i>	DNA mismatch repair protein MutS	804				OM	C				2.2.03
2710	<i>nadE</i>	glutamine-dependent NAD(+) synthetase*	513				OM	U			6.3.5.1	3.2.11
2720	<i>hemB</i>	delta-aminolevulinic acid dehydratase	329				C	U			4.2.1.24	3.2.06
2730		unknown	912				OM	C				0.0.0
2740		integral membrane transport protein*	426	12		12 th	IM	IM				1.5.5
2750		membrane protein*	527	1	signal	2 th	E	IM				4.1.1
2760		membrane protein*	519	1	signal	2 th	OM	U				4.1.1
2770		membrane protein*	526	1	signal	2 th	C	U				4.1.1
2780		membrane protein*	524	1	signal	2 th	C	U		21 bp x 2.0		4.1.1
2790		integral membrane protein*	653	2 [†]		2 th	C	OM				4.1.1

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
2800		membrane protein*	520	1	signal	2 th	C	U		15 bp x 2.0		4.1.1
2810		integral membrane transport protein*	417	12		12 th	IM	IM				1.5.5
2820		integral membrane transport protein*	415	12		12 th	IM	IM				1.5.5
2830	<i>ssb</i>	single-strand DNA binding protein	156				C	C				2.2.03
2840	<i>matA</i>	malonyl-CoA decarboxylase*	460				C	C			4.1.1.9	3.6.1
2850	<i>gatB</i>	aspartyl/glutamyl-tRNA amidotransferase subunit B	481				C	C			6.3.5.-	2.2.01
2860	<i>fabI</i>	enoyl-[acyl-carrier-protein] reductase [NADH]	273				C	U			1.3.1.9	3.6.1
2870	<i>dnaA</i>	chromosomal replication initiator protein DnaA	464				C	C				2.2.03
2900		integral membrane protein*	331	2		2 th	C	C				4.1.1
2910	<i>nadD</i>	nicotinate-nucleotide adenyltransferase*	194		signal		OM	U			2.7.7.18	3.2.11
2920	<i>pdxJ</i>	pyridoxal phosphate biosynthetic protein PdxJ	238				C	C				3.2.12
2930	<i>hupB</i>	DNA-binding protein HU-beta*	94				C	U				2.2.03
2940	<i>holB</i>	DNA polymerase III, delta prime subunit [†]	296				C	C			2.7.7.7	2.2.03
2950		conserved hypothetical protein	199				C	U		144 bp x 6.0 (N-terminus)		0.0.2
2960	<i>coaE</i>	dephospho-CoA kinase*	201				C	U			2.7.1.24	3.2.17
2970	<i>thiC</i>	thiamine biosynthesis protein ThiC	555				C	U				3.2.14
2980		unknown	186				E	U				7.0.0
2990	<i>rpoZ</i>	DNA-directed RNA polymerase omega chain*	132				C	C			2.7.7.6	2.2.11
3000		unknown	123				C	U				7.0.0
3010	<i>leuS</i>	leucyl-tRNA synthetase	830				C	C			6.1.1.4	2.2.01
3030		deoxyribonuclease [†]	261				C	C			3.1.21.-	7.0.0
3040	<i>pyrF</i>	orotidine 5'-phosphate decarboxylase	231				C	IM			4.1.1.23	3.7.2
3050	<i>surE</i>	acid phosphatase SurE	252				E	U			3.1.3.2	1.4.0
3060	<i>ccmE</i>	cytochrome c-type biogenesis protein CcmE	134	1 [†]		signal	P	U				2.2.13
3070	<i>nuoC</i>	NADH-quinone oxidoreductase chain C*	191				C	C			1.6.99.5	3.5.3
3090	<i>nuoB</i>	NADH-quinone oxidoreductase chain B	172				P	U			1.6.99.5	3.5.3
3100	<i>nuoA</i>	NADH-quinone oxidoreductase chain A*	123	3		3 th	IM	IM			1.6.99.5	3.5.3
3110	<i>uvrA</i>	uvrABC system protein A	959				OM	OM				2.2.03
3120		unknown	174				C	U				7.0.0
3130	<i>ribH</i>	6,7-dimethyl-8-ribityllumazine synthase*	149				C	U			2.5.1.9	3.2.13
3140		integral membrane protein*	575	6 [†]		6 th	IM	IM				4.1.1
3150		integral membrane transport protein*	461	11		10 th	IM	IM				1.5.5
3160	<i>pssA</i>	CDP-diacylglycerol--serine O-phosphatidyltransferase*	260	7			IM	IM			2.7.8.-	2.2.07
3170	<i>psd</i>	phosphatidylserine decarboxylase proenzyme*	227	1		1 th	IM	U			4.1.1.65	2.2.07
3180		unknown	1134				C	U		182 bp x 4.4 (C-terminus)		7.0.0
3190	<i>efp</i>	elongation factor P*	189				C	C				2.2.10
3200	<i>suhB</i>	inositol-1-monophosphatase*	256				OM	U			3.1.3.25	6.1.3

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
3210	<i>rluC</i>	ribosomal large subunitPseudouridine synthase C*	305				C	C			4.2.1.70	2.2.11
3220		response regulator component of a two-component regulatory system [†]	461				C	C				6.0.0
3221		unknown	93				C	U				0.0.0
3230		NAD-glutamate dehydrogenase [†]	1589				C	OM			1.4.1.-	3.3.20
3240		integral membrane protein*	210	4		4 th	IM	IM				4.1.1
3250	<i>cysS</i>	cysteinyl-tRNA synthetase	457				C	C			6.1.1.16	2.2.01
3260		unknown	648				OM	U				0.0.0
3270	<i>nrdB</i>	ribonucleoside-diphosphate reductase beta chain*	324	1		1 th	C	U			1.17.4.1	3.7.0
3280		conserved hypothetical protein	94				P	U				0.0.1
3290		unknown	194				E	U				0.0.0
3300		conserved hypothetical protein	339				C	U				0.0.2
3310	<i>dnaG</i>	DNA primase*	592				C	C			2.7.7.-	2.2.03
3320	<i>rpoD</i>	RNA polymerase sigma-70 factor	622				C	C	+++			2.2.11
3330		conserved hypothetical protein	317				C	U				0.0.2
3340	<i>ispE</i>	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase*	281				C	U			2.7.1.148	3.2.08
3350	<i>cutA</i>	periplasmic divalent cation tolerance protein CutA*	109				C	C				1.6.4
3360		two component sensor kinase*	828	3		3 th	C	IM				6.1.0
3370	<i>mdmC</i>	O-methyltransferase*	218				E	U			2.1.1.-	2.2.0
3380		unknown	94				C	U				0.0.0
3390		conserved hypothetical protein	467			signal	IM	U				0.0.2
3400	<i>topA</i>	DNA topoisomerase I	819				OM	U			5.99.1.2	2.2.03
3410		unknown	119				C	U				0.0.0
3420		conserved hypothetical protein	153				C	U				0.0.1
3430	<i>acpS</i>	holo-[acyl-carrier-protein] synthase*	123				C	U			2.7.8.7	3.6.1
3440	<i>proS</i>	prolyl-tRNA synthetase	426				C	C			6.1.1.15	2.2.01
3450		exported protein*	327		signal	signal	C	U				4.1.0
3460	<i>coaD</i>	phosphopantetheine adenylyltransferase*	165				C	U			2.7.7.3	3.2.17
3470	<i>trxB</i>	thioredoxin reductase	318			signal	C	U			1.8.1.9	3.2.15
3480		peroxiredoxin*	205				C	C				1.4.1
3490	<i>aatA</i>	aspartate aminotransferase A	399				OM	U			2.6.1.1	3.1.07
3500	<i>ppiD</i>	peptidyl-prolyl cis-trans isomerase D*	630	1 [†]		signal	OM	OM	+		5.2.1.8	1.3.1
3510		glycoprotease [†]	193				C	U			3.4.-.-	2.1.4
3520	<i>truB</i>	tRNAPseudouridine synthase B*	296				C	C			4.2.1.70	2.2.11
3530	<i>rpsO</i>	30S ribosomal protein S15	93				C	U				4.2.2
3540	<i>pnp</i>	polyribonucleotide nucleotidyltransferase	789				OM	C			2.7.7.8	2.1.2
3550		conserved hypothetical protein	265				IM	IM				0.0.2
3560	<i>lepA</i>	GTP-binding protein LepA	598				C	C				2.2.10
3570		integral membrane protein*	376	2		2 th	E	U		12 bp x 3.2		4.1.1

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
3580		integral membrane protein*	188	2 [‡]		signal, 1 th	OM	U				4.1.1
3590		integral membrane protein*	389	2		2 th	E	U		45 bp x 7.4, 42 bp x 2.1		4.1.1
3600		integral membrane protein*	585	2 [‡]		2 th	OM	IM		12 bp x 16.7		4.1.1
3610		membrane protein*	513	1	signal	2 th	C	U				4.1.1
3620		integral membrane protein*	537	2 [‡]		2 th	OM	U				4.1.1
3630		membrane protein*	519	1	signal	signal, 1 th	OM	U				4.1.1
3640		unknown	111				C	U				0.0.0
3650	<i>prfB</i>	peptide chain release factor 2	367				OM	C				2.2.10
3660		conserved hypothetical protein	249				C	U				0.0.2
3670	<i>gatA</i>	glutamyl-tRNA(Gln) amidotransferase subunit A	487				OM	U			6.3.5.7	2.2.01
3680	<i>folC</i>	folypolyglutamate synthase/dihydrofolate synthase*	431				C	U			6.3.2.17, 6.3.2.12	3.2.05
3690	<i>hemC</i>	porphobilinogen deaminase	299				C	U			2.5.1.61	3.2.06
3700	<i>typA</i>	GTP-binding protein TypA/BipA*	612				C	C				7.0.0
3701		unknown	106				C	U				0.0.0
3710	<i>nuoI</i>	NADH-quinone oxidoreductase chain I	168	1		1 th	C	C			1.6.99.5	3.5.3
3720	<i>sipF</i>	prokaryotic type I signal peptidase	238	1		signal	C	IM			3.4.21.89	2.2.10
3730		unknown	153				C	U		27 bp x 2.2		7.0.0
3740		metal dependent phosphohydrolase*	403				C	C				7.0.0
3750		unknown	1674				OM	C		27 bp x 8.3, 144 bp x 3.9		7.0.0
3760		integral membrane protein*	308	6		6 th	IM	IM				4.1.1
3770	<i>argG</i>	argininosuccinate synthase	394				C	C			6.3.4.5	3.1.02
3780		exported protein*	223		signal	signal	OM	IM				4.1.0
3790		exported protein*	235		signal	signal	OM	OM				4.1.0
3800	<i>argJ</i>	arginine biosynthesis bifunctional protein ArgJ	419				E	C			2.3.1.1, 2.3.1.35	3.1.02
3810	<i>exoA</i>	exodeoxyribonuclease*	278				C	C			3.1.11.2	2.2.03
3820		integral membrane protein*	260	6		6 th	IM	IM				4.1.1
3830		integral membrane protein*	276	6		6 th	IM	IM				4.1.1
3840	<i>fabG</i>	3-oxoacyl-[acyl carrier protein] reductase	245				C	C			1.1.1.100	3.6.1
3850	<i>putA</i>	proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase	1043			1 th	C	C			1.5.99.8, 1.5.1.12	3.1.16
3860		membrane protein*	171	1 [‡]		1 th	E	U				4.1.1
3870	<i>bioA</i>	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	425				C	U			2.6.1.62	3.2.02
3880		conserved hypothetical protein	471				OM	U				0.0.2
3890		unknown	126				P	U				0.0.0
3900		unknown	189				C	U				0.0.0
3910		unknown	129				C	U				0.0.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
3920		unknown	136				C	U				0.0.0
3930		unknown	188				C	C				0.0.0
3940		unknown	115				C	C				0.0.0
3950	<i>rpmJ</i>	50S ribosomal protein L36	42				C	U				4.2.2
3960	<i>rpoH</i>	RNA polymerase sigma-32 factor	296				C	C	+			2.2.11
3970		unknown	198				C	C				7.0.0.
3980		unknown	3002				OM	E		144 bp x 2.7, 36 bp x 2.7, 93 bp x 9.2		7.0.0
3990	<i>atpG</i>	ATP synthase gamma chain	283			1 th	OM	U			3.6.3.14	3.5.9
4000	<i>folE</i>	GTP cyclohydrolase I	190				C	IM			3.5.4.16	3.2.05
4010	<i>pmbA</i>	PmbA protein*	455				OM	C				1.3.1
4020		pyridine nucleotide-disulphide oxidoreductase*	337	1			OM	C				3.3.00
4030	<i>ksgA</i>	dimethyladenosine transferase	262	1			C	U			2.1.1.-	2.2.11
4040	<i>tpiA</i>	triosephosphate isomerase	240				OM	U			5.3.1.1	3.3.04
4050		exported protein*	326		signal	signal	C	IM				4.1.0
4060	<i>gcp</i>	o-sialoglycoprotein endopeptidase	348				C	E			3.4.24.57	2.1.4
4070		integral membrane protein*	193	3		3 th	IM	IM				4.1.1
4080	<i>folB</i>	dihydroneopterin aldolase [†]	115				C	U			4.1.2.25	3.2.05
4090	<i>mdh</i>	malate dehydrogenase	314				C	C			1.1.1.37	3.5.8
4100	<i>rpiB</i>	ribose 5-phosphate isomerase B	146				C	C			5.3.1.6	3.3.09
4110	<i>ubiG</i>	3-demethylubiquinone-9 3-methyltransferase*	241				C	C			2.1.1.64	3.2.08
4120		conserved hypothetical protein	156				C	U				0.0.2
4130		conserved hypothetical protein	240				P	U				0.0.2
4140		unknown	522				OM	U				7.0.0
4150	<i>iscS</i>	cysteine desulfurase	413				C	C			4.4.1.-	2.2.11
4160		NifU-like protein*	137				P	U				3.3.19
4170		conserved hypothetical protein	149				C	U				0.0.2
4180	<i>hscB</i>	co-chaperone protein HscB [†]	145				C	U				1.3.1
4190	<i>hscA</i>	chaperone protein HscA	616				C	C				1.3.1
4200	<i>fdxB</i>	ferredoxin, 2FE-2S	122				C	C				3.5.3
4210		membrane protein*	356	1 [†]		signal	OM	U				4.1.1
4211		cytochrome c-type biogenesis protein [†]	125	1		1 th	C	U				2.2.13
4220	<i>lysS</i>	lysyl-tRNA synthetase	512				C	C		21 bp x 2.0	6.1.1.6	2.2.01
4230		integral membrane protein*	135	2		signal	C	U				4.1.1
4240	<i>truA</i>	tRNAPseudouridine synthase A	246				C	U			4.2.1.70	2.2.01
4250	<i>pyrB</i>	aspartate carbamoyltransferase	306				C	C			2.1.3.2	3.7.2
4260	<i>gyrB</i>	DNA gyrase subunit B	798				C	U			5.99.1.3	2.2.03
4261		unknown	84				P	U				0.0.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
4270	<i>nuoG</i>	NADH-quinone oxidoreductase chain G	684				C	C			1.6.99.5	3.5.3
4280	<i>nuoH</i>	NADH-quinone oxidoreductase chain H	367	8		8 th	IM	IM			1.6.99.5	3.5.3
4310	<i>gltX2</i>	glutamyl-tRNA synthetase 2	470				C	C			6.1.1.17	2.2.01
4320		unknown	425				C	C				0.0.0
4330	<i>mutM</i>	formamidopyrimidine-DNA glycosylase	269				C	U			3.2.2.23	2.2.03
4340		unknown	392				OM	U				0.0.0
4350		unknown	409				OM	U				0.0.0
4360		unknown	157				OM	U				0.0.0
4370	<i>miaA</i>	tRNA delta(2)-isopentenylpyrophosphate transferase*	300				C	U			2.5.1.8	2.2.11
4390		unknown	240				C	C				0.0.0
4400		unknown	994				C	U				0.0.0
4410		type IV secretion system protein [†]	232	1		1 th	OM	OM				1.5.6
4420	<i>nuoD</i>	NADH-quinone oxidoreductase chain D	393				C	C			1.6.99.5	3.5.3
4430	<i>nuoE</i>	NADH-quinone oxidoreductase chain E	183				C	C			1.6.99.5	3.5.3
4440		integral membrane protein*	195	#4		4 th	C	IM				4.1.1
4450		unknown	280				OM	U				0.0.0
4460	<i>pccB</i>	propionyl-CoA carboxylase beta chain	510				C	U			6.4.1.3	3.4.4
4470		exported protein*	385		signal	signal	OM	OM				4.1.0
4480	<i>argB</i>	acetylglutamate kinase	305				C	U			2.7.2.8	3.1.02
4490	<i>engB</i>	GTP binding protein EngB*	200				C	U				1.7.1
4500	<i>prfA</i>	peptide release factor 1	359				C	C				2.2.10
4510		sodium:dicarboxylate symporter (glutamate)*	402	8		8 th	IM	IM				1.5.1
4520	<i>rmuC</i>	DNA recombination protein RmuC	436	1		signal	C	U				2.2.03
4530		unknown	199				E	U		22 bp x 2.0		0.0.0
4540	<i>serS</i>	seryl-tRNA synthetase	427				C	C			6.1.1.11	2.2.01
4550	<i>hemF</i>	coproporphyrinogen III oxidase	288				E	U			1.3.3.3	3.2.06
4560		conserved hypothetical protein	95				C	C				0.0.1
4570	<i>tal</i>	transaldolase*	220				C	U			2.2.1.2	3.3.09
4580	<i>atpC</i>	ATP synthase epsilon chain [†]	134				C	C			3.6.3.14	3.5.9
4590	<i>atpD</i>	ATP synthase beta chain	504				C	U			3.6.3.14	3.5.9
4600		magnesium transporter*	456	4		5 th	IM	IM				1.5.2
4610		membrane protein*	124	1 [‡]		1 th	C	C				4.1.1
4620		membrane protein*	134	1 [‡]		1 th	IM	U				4.1.1
4630		membrane protein*	125	1 [‡]		1 th	P	U				4.1.1
4640		membrane protein*	123	1		1 th	P	U				4.1.1
4650		unknown	771				C	C				0.0.0
4660	<i>clpA</i>	ATP-dependent Clp protease, ATP-binding subunit	764				C	C			3.4.21.92	2.1.4
4670		conserved hypothetical integral membrane protein	235	7		7 th	IM	IM				4.1.1
4680	<i>rbfA</i>	ribosome-binding factor A*	115				C	C				2.2.11

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
4690	<i>infB</i>	translation initiation factor IF-2	856				C	C				2.2.10
4700	<i>nusA</i>	N utilization substance protein A*	517				C	C				2.2.11
4710		integral membrane transport protein*	1039	12		12 th	IM	IM				1.5.5
4720	<i>tatC</i>	Sec-independent protein translocase protein TatC	250	5		6 th	IM	IM				1.5.5
4730	<i>ispG</i>	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase*	422				C	U		247 bp x 2.1 (N-terminus)	1.17.4.3	3.2.08
4740		exported protein*	639		signal	signal	E	OM		138 bp x 6.9		4.1.0
4750	<i>dxr</i>	1-deoxy-D-xylulose 5-phosphate reductoisomerase	387	1		1 th	IM	U			1.1.1.267	3.2.08
4760	<i>nuoN</i>	NADH-quinone oxidoreductase chain N	474	13		14 th	IM	IM			1.6.99.5	3.5.3
4770	<i>nuoM</i>	NADH-quinone oxidoreductase chain M	486	14		14 th, signal	IM	IM			1.6.99.5	3.5.3
4780	<i>nuoL</i>	NADH-quinone oxidoreductase chain L	622	16		17 th, signal	IM	IM	+		1.6.99.5	3.5.3
4790	<i>nuoK</i>	NADH-quinone oxidoreductase chain K	108	3		3 th	IM	IM			1.6.99.5	3.5.3
4800	<i>nuoJ</i>	NADH-quinone oxidoreductase chain J	200	5		5 th	IM	IM			1.6.99.5	3.5.3
4810	<i>nuoF</i>	NADH-quinone oxidoreductase chain F	425			1 th	C	C			1.6.99.5	3.5.3
4820	<i>rpmA</i>	50S ribosomal protein L27	88				C	U				4.2.2
4830	<i>rplU</i>	50S ribosomal protein L21	102				C	U				4.2.2
4840	<i>eno</i>	enolase	421				C	C			4.2.1.11	3.3.15
4850		conserved hypothetical GTP-binding protein	340				C	C		9 bp x 3.0		0.0.2
4860	<i>mraW</i>	S-adenosyl-methyltransferase MraW*	301				OM	C			2.1.1.-	7.0.0
4870	<i>ileS</i>	isoleucyl-tRNA synthetase	1104				C	C			6.1.1.5	2.2.01
4880		bacterioferritin comigratory protein [†]	147				C	U				7.0.0
4890		conserved hypothetical protein	238				C	C				0.0.2
4900		unknown	173				C	C				7.0.0
4910	<i>argS</i>	arginyl-tRNA synthetase	576			1 th	C	C			6.1.1.19	2.2.01
4920	<i>recO</i>	DNA repair protein RecO [†]	244				C	U				2.2.03
4930		unknown	186				E	U				0.0.0
4950		competence protein [†]	492				C	C				7.0.0
4960		integral membrane protein*	129	2		2 th	P	U				4.1.1
4970	<i>rbn</i>	tRNA processing ribonuclease BN [†]	277	5		6 th	IM	IM				2.2.12
4980	<i>thiL</i>	thiamine-monophosphate kinase*	316				C	IM			2.7.4.16	3.2.14
4990	<i>dnaQ</i>	DNA polymerase III, epsilon subunit	242				C	C			2.7.7.7	2.2.03
5000		exported protein*	490		signal	signal	OM	OM				4.1.0
5010		exported protein*	564		signal	signal	OM	OM		24 bp x 2.1		4.1.0
5020	<i>petC</i>	cytochrome c1 precursor	252	1	signal	1 th, signal	P	U				3.5.3
5030	<i>petB</i>	cytochrome b	408	9		9 th	IM	IM		20 bp x 2.0		3.5.3
5040	<i>petA</i>	ubiquinol-cytochrome c reductase iron-sulphur subunit	187	1		1 th	C	U			1.10.2.2	3.5.3
5050		integral membrane protein*	290	5		5 th	IM	IM				4.1.1
5060		ABC transporter, membrane-spanning protein*	266	7		8 th	IM	IM				1.5.7

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
5070		conserved hypothetical protein	120				C	C				0.0.2
5080	<i>tsf</i>	elongation factor Ts	288				C	C				2.2.10
5090	<i>rpsB</i>	30S ribosomal protein S2	286				C	U				4.2.2
5100	<i>maf</i>	septum formation protein Maf [†]	192				OM	U				7.0.0
5110	<i>infA</i>	translation initiation factor IF-1*	82				C	C				2.2.10
5120		secretion protein [†]	363	1		1 th	OM	U				7.0.0
5130		dihydrolipoamide dehydrogenase, E3 component of pyruvate or 2-oxoglutarate dehydrogenase complex*	465				C	C			1.8.1.4	3.5.0
5140		exported protein*	389		signal	signal	OM	U				4.1.0
5150	<i>gpml</i>	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	501				C	C			5.4.2.1	3.3.15
5160	<i>engA</i>	GTP binding protein EngA*	439				C	U				7.0.0
5170	<i>carA</i>	carbamoyl-phosphate synthase small chain	374				E	U			6.3.5.5	3.1.02
5180	<i>ispH</i>	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	328				C	U			1.17.1.2	3.2.08
5190	<i>dut</i>	deoxyuridine 5'-triphosphate nucleotidohydrolase*	124				C	U			3.6.1.23	3.3.11
5200		conserved hypothetical protein	203				C	U				0.0.1
5210		type IV secretion system protein [†]	2455	4	signal	6 th, signal	OM	OM		261 bp x 5.1, 222 bp x 3.9, 180 bp x 1.9		1.5.6
5220		type IV secretion system protein [†]	1529	6	signal	5 th, signal	OM	OM		216 bp x 2.4, 15 bp x 2.3		1.5.6
5230		type IV secretion system protein [†]	911	7	signal	7 th, signal	OM	IM				1.5.6
5240	<i>virB6</i>	type IV secretion system protein VirB6	821	8	signal	9 th, signal	IM	IM				1.5.6
5250	<i>virB4</i>	type IV secretion system protein VirB4	800				C	U				1.5.6
5260	<i>virB3</i>	type IV secretion system protein VirB3	97	2		2 th	IM	U				1.5.6
5270	<i>sodB</i>	superoxide dismutase [Fe]	210				E	U			1.15.1.1	1.4.2
5280		ABC transporter, membrane-spanning protein*	420	6		6 th	IM	IM				1.5.7
5290	<i>lipA</i>	lipoic acid synthetase	292				C	C				3.2.07
5300		unknown	464				OM	U				0.0.0
5310		integral membrane protein*	1392	2		3 th	C	C				4.1.1
5320	<i>bccA</i>	acetyl-/propionyl-coenzyme A carboxylase alpha chain*	660				C	C		14 bp x 2.1	6.3.4.14	3.6.1
5330	<i>rluD</i>	ribosomal large subunitPseudouridine synthase D	324				C	U			4.2.1.70	2.2.11
5340	<i>lysA</i>	diaminopimelate decarboxylase*	420				C	U			4.1.1.20	3.1.13
5350	<i>rpmB</i>	50S ribosomal protein L28	100				C	U				4.2.2
5360	<i>priA</i>	primosomal protein N'	659				OM	U				2.2.03
5370		exported protein*	325		signal	signal	C	U				4.1.0
5380	<i>hemD</i>	uroporphyrinogen-III synthase*	242				C	U			4.2.1.75	3.2.06
5390		aminopeptidase [†]	572				C	U				7.0.0
5400		unknown	173			1 th	OM	U				7.0.0
5410		conserved hypothetical protein	275				C	C				0.0.2
5420	<i>era</i>	GTP-binding protein ERA*	296				C	U				7.0.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
5430	<i>ffh</i>	signal recognition particle protein	450				C	C				1.5.5
5440		NADH-quinone oxidoreductase subunit*	528	15		16 th	IM	IM			1.6.99.5	3.5.3
5450		unknown	264				OM	U				0.0.0
5460		unknown	258				C	U				0.0.0
5470		membrane protein*	158	1		1 th	C	U				4.1.1
5480		membrane protein*	111	1 [‡]		signal	C	IM				4.1.1
5490		conserved hypothetical protein	153				C	U				0.0.2
5500	<i>dnaK</i>	chaperone protein DnaK	645				C	U				1.3.1
5510		ribonuclease*	610				OM	U			3.1.4.-	2.1.2
5520		integral membrane protein*	111	3		3 th	IM	IM				4.1.1
5530		Na ⁺ /H ⁺ antiporter subunit*	139	4		4 th	IM	IM				1.5.2
5540		integral membrane protein*	181	6		6 th	IM	IM				4.1.1
5550		Na ⁺ /H ⁺ antiporter subunit*	99	3		3 th	IM	IM				1.5.2
5560		integral membrane protein*	88	3		2 th, signal	IM	IM				4.1.1
5570		unknown	552				OM	U		183 bp x 3.0		0.0.0
5580		unknown	344				E	U				0.0.0
5590		unknown	213				C	C				0.0.0
5600	<i>tkt</i>	transketolase	671				C	U			2.2.1.1	3.3.09
5610		carboxypeptidase [†]	491				C	U				2.1.4
5620		unknown	217			signal	P	OM				7.0.0
5630	<i>purA</i>	adenylosuccinate synthetase	430				C	C			6.3.4.4	3.7.1
5640		Holliday junction resolvase [†]	156				C	U			3.1.-.-	2.2.03
5650	<i>nrdA</i>	ribonucleoside-diphosphate reductase alpha chain*	595				C	U			1.17.4.1	3.7.0
5660	<i>ispA</i>	geranyltranstransferase*	276				C	C			2.5.1.10	3.2.08
5670		membrane protein*	317	1 [‡]		signal	C	U				4.1.1
5680	<i>thiO</i>	thiamine biosynthesis oxidoreductase*	354		signal		C	C				3.2.14
5690		deaminase [†]	145				C	U			3.5.4.-	7.0.0
5700		membrane protein*	142	1		1 th	C	U				4.1.1
5710	<i>dnaB</i>	replicative DNA helicase	486				C	C	+		3.6.1.-	2.2.03
5720	<i>fabH</i>	3-oxoacyl-[acyl-carrier-protein] synthase III	319	1		1 th	IM	U			2.3.1.41	3.6.1
5730	<i>plsX</i>	fatty acid/phospholipid synthesis protein	336				C	C				2.2.07
5740	<i>rpmF</i>	50S ribosomal protein L32	60				C	U				4.2.2
5750	<i>tgt</i>	queuine tRNA-ribosyltransferase	378				C	U			2.4.2.29	2.2.12
5760	<i>pstB</i>	phosphate ABC transporter, ATP-binding protein*	253				C	U			3.6.3.27	1.5.7
5770	<i>dapB</i>	dihydrodipicolinate reductase	264				C	U			1.3.1.26	3.1.13
5780		monooxygenase*	164				C	U				7.0.0
5790	<i>ubiA</i>	4-hydroxybenzoate octaprenyltransferase	295	8		8 th	IM	IM			2.5.1.-	3.2.08
5791	<i>rpmH</i>	50S ribosomal protein L34	44				C	U				4.2.2
5800	<i>rnpA</i>	ribonuclease P protein component*	122				C	U			3.1.26.5	2.2.11

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
5810		integral membrane transport protein*	437	11		11 th	IM	IM				1.5.5
5820		competence protein [†]	650	12		11 th	IM	IM		142 bp x 10 (C-terminus)		1.5.5
5830	<i>pheT</i>	phenylalanyl-tRNA synthetase beta chain	789				C	C			6.1.1.20	2.2.01
5840	<i>rplQ</i>	50S ribosomal protein L17	128				C	C				4.2.2
5850	<i>rpoA</i>	DNA-directed RNA polymerase alpha chain	358				C	C			2.7.7.6	2.2.11
5860	<i>rpsK</i>	30S ribosomal protein S11	127				P	U				4.2.2
5870	<i>rpsM</i>	30S ribosomal protein S13	123				C	C				4.2.2
5880	<i>adk</i>	adenylate kinase	220				C	C			2.7.4.3	3.7.1
5890	<i>secY</i>	preprotein translocase secY subunit	432	11		11 th	IM	IM				1.5.5
5900	<i>rplO</i>	50S ribosomal protein L15	157				C	U				4.2.2
5910	<i>rpsE</i>	30S ribosomal protein S5	174				C	C				4.2.2
5920	<i>rplR</i>	50S ribosomal protein L18	120				C	C				4.2.2
5930	<i>rplF</i>	50S ribosomal protein L6	178				OM	U				4.2.2
5940	<i>rpsH</i>	30S ribosomal protein S8	132				C	C				4.2.2
5950	<i>rpsN</i>	30S ribosomal protein S14	101				C	U				4.2.2
5960	<i>rplE</i>	50S ribosomal protein L5	177				C	U				4.2.2
5970	<i>rplX</i>	50S ribosomal protein L24	109				P	U				4.2.2
5980	<i>rplN</i>	50S ribosomal protein L14	119				C	U				4.2.2
5990	<i>rpsQ</i>	30S ribosomal protein S17	74				C	U				4.2.2
5991	<i>rpmC</i>	50S ribosomal protein L29	67				C	C				4.2.2
6000	<i>rplP</i>	50S ribosomal protein L16	136				P	U				4.2.2
6010	<i>rpsC</i>	30S ribosomal protein S3	211				C	U				4.2.2
6020	<i>rplV</i>	50S ribosomal protein L22	114				C	U				4.2.2
6030	<i>rpsS</i>	30S ribosomal protein S19	93				P	U				4.2.2
6040	<i>rplB</i>	50S ribosomal protein L2	276				E	U				4.2.2
6050	<i>rplW</i>	50S ribosomal protein L23	96				C	U				4.2.2
6060	<i>rplD</i>	50S ribosomal protein L4	205				C	U				4.2.2
6070	<i>rplC</i>	50S ribosomal protein L3	231				P	U				4.2.2
6080	<i>rpsJ</i>	30S ribosomal protein S10	111				C	U				4.2.2
6090	<i>tufB</i>	elongation factor Tu-B	395				C	C				2.2.10
6100		tRNA/rRNA methyltransferase*	249				C	U			2.1.1.-	2.2.11
6110	<i>cmk</i>	cytidylate kinase*	212				C	U			2.7.4.14	3.7.2
6120	<i>rpsA</i>	30S ribosomal protein S1	565				OM	C				4.2.2
6130		peptidase*	289	1		1 th	OM	U			3.4.21.-	2.1.4
6140	<i>ihfB</i>	integration host factor beta subunit [†]	87				C	U				2.2.03
6150		unknown	97				C	C				0.0.0
6160		unknown	121				C	C				0.0.0
6170		integral membrane protein*	325	5		4 th	IM	IM				4.1.1

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
6180	<i>hemH</i>	ferrochelatase	313			signal	C	C			4.99.1.1	3.2.06
6190		ATPase*	357			1 th	C	U				7.0.0
6200		unknown	101				C	U				0.0.0
6210		integral membrane protein*	285	3		3 th	IM	IM				4.1.1
6220		unknown	125				C	U				7.0.0
6230		exported protein*	134		signal	signal	P	U				4.1.0
6240		membrane protein*	81	1 [†]		signal	P	U				4.1.1
6250	<i>recB</i>	exodeoxyribonuclease V beta chain*	857				C	U	+	132 bp x 5.4 (C-terminus)	3.1.11.5	2.2.03
6260	<i>qor</i>	quinone oxidoreductase*	324				C	U			1.6.5.5	3.5.3
6270		ABC transporter, ATP-binding protein*	593	5		5 th	IM	IM				1.5.7
6280	<i>folP1</i>	dihydropteroate synthase 1*	274				C	U			2.5.1.15	3.2.05
6290	<i>folP2</i>	dihydropteroate synthase 2*	283				C	C			2.5.1.15	3.2.05
6300		integral membrane protein*	352	2		2 th	C	U				4.1.1
6310	<i>carB</i>	carbamoyl-phosphate synthase, large subunit	1075				C	U			6.3.5.5	3.1.02
6320		unknown	105				C	U				0.0.0
6330	<i>fumC</i>	fumarate hydratase class II	461				C	C			4.2.1.2	3.5.8
6350	<i>pyrC</i>	dihydroorotase	449				C	U			3.5.2.3	3.7.2
6360	<i>lipB</i>	lipoate-protein ligase B	208				C	C			6.-.-.	2.2.0
6370	<i>purN</i>	phosphoribosylglycinamide formyltransferase	212				IM	U			2.1.2.2	3.7.1
6380	<i>pepA</i>	cytosol aminopeptidase	500			1 th	OM	U			3.4.11.1	2.1.4
6390		dioxygenase*	244	1		1 th	E	U			1.3.11.-	7.0.0
6400	<i>clpB</i>	heat shock protein ClpB	859				C	C				1.3.1
6410		oxidoreductase*	249				OM	C			1.-.-.	7.0.0
6420	<i>groEL</i>	60 kDa chaperonin GroEL	551				C	U				1.3.1
6430	<i>groES</i>	10 kDa chaperonin GroES	94				C	C				1.3.1
6440	<i>radC</i>	DNA repair protein RadC	230				C	U				2.2.03
6450	<i>purQ</i>	phosphoribosylformylglycinamide synthase I [†]	265				C	U			6.3.5.3	3.7.1
6460	<i>gidA</i>	glucose inhibited division protein A	625				OM	U				1.7.1
6470	<i>glpX</i>	fructose-1,6-bisphosphatase class II GlpX	306				C	C			3.1.3.11	3.3.04
6480		peptidase*	204			signal	C	U				7.0.0
6490		unknown	151				C	C				0.0.0
6500	<i>bioB</i>	biotin synthase	322				C	C			2.8.1.6	3.2.02
6510	<i>purL</i>	phosphoribosylformylglycinamide synthase II*	1010				C	U		185 bp x 6.2 (C-terminus)	6.3.5.3	3.7.1
6520	<i>folK</i>	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase*	174				C	U		124 bp x 4.4 (C-terminus)	2.7.6.3	3.2.05
6530		unknown	210			signal	E	U				0.0.0
6540		zinc metallopeptidase*	433	2 [†]		signal	IM	IM				7.0.0
6550		conserved hypothetical protein	121				C	U				0.0.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
6560		unknown	295				C	U				0.0.0
6570		unknown	212				C	U				0.0.0
6580	<i>purM</i>	phosphoribosylformylglycinamide cyclo-ligase	342				C	U			6.3.3.1	3.7.1
6590		integrase/recombinase XerD or XerC*	312				C	U				2.2.03
6600	<i>gpsA</i>	glycerol-3-phosphate dehydrogenase [NAD(P)+]	327			signal	C	IM			1.1.1.94	2.2.07
6610		response regulator component of a two-component regulatory system*	471				C	C	+			6.1.2
6620	<i>ftsQ</i>	cell division protein FtsQ*	271	1		1 th	C	C				1.7.1
6640	<i>gshB</i>	glutathione synthetase	312				C	C			6.3.2.3	3.2.18
6650		exported protein*	200		signal	signal	IM	P				4.1.0
6660	<i>aspS</i>	aspartyl-tRNA synthetase	590				C	C			6.1.1.12	2.2.01
6670		haloacid dehalogenase-like hydrolase*	210				E	U			3.1.3.-	7.0.0
6680		integral membrane protein*	170	3		3 th	IM	IM				4.1.1
6690	<i>ppdK</i>	pyruvate phosphate dikinase	873				C	C			2.7.9.1	3.3.15
6700		NADH-quinone oxidoreductase subunit*	492	13		14 th	IM	IM			1.6.99.5	3.5.3
6710		conserved hypothetical protein	258			1 th	C	U				0.0.2
6720		c-type cytochrome*	174	1		1 th	P	IM				3.5.3
6730	<i>folD</i>	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	300				C	U			1.5.1.5 3.5.4.9	3.2.05
6740	<i>gmk</i>	guanylate kinase	209				C	C			2.7.4.8	3.7.1
6750	<i>ccmC</i>	heme exporter protein C	234	6		6 th	IM	IM				1.5.7
6760	<i>ruvA</i>	Holliday junction DNA helicase RuvA*	191				C	U				2.2.03
6770	<i>ruvB</i>	Holliday junction DNA helicase RuvB	331				C	C				2.2.03
6780	<i>bcr</i>	bicyclomycin resistance protein*	398	12		12 th	IM	IM				1.5.5
6790	<i>thyX</i>	thymidylate synthase complementing protein*	285				C	U			2.1.1.148	7.0.0
6800	<i>sdhB</i>	succinate dehydrogenase iron-sulfur subunit	258				C	U			1.3.99.1	3.5.8
6810	<i>sdhA</i>	succinate dehydrogenase flavoprotein subunit	598				C	U			1.3.99.1	3.5.8
6820		ABC transporter, ATP-binding and membrane-spanning protein*	583	5		6 th	OM	IM				1.5.7
6830		unknown	109				C	U				0.0.0
6840	<i>glyA</i>	serine hydroxymethyltransferase	421				C	C			2.1.2.1	3.1.09
6850	<i>rplI</i>	50S ribosomal protein L9	207				C	C				4.2.2
6860	<i>rpsR</i>	30S ribosomal protein S18	95				C	U				4.2.2
6870	<i>rpsF</i>	30S ribosomal protein S6	109				C	U				4.2.2
6880		integral membrane protein*	203	3		3 th	IM	IM				4.1.1
6890		integral membrane protein*	881	20		22 th	IM	IM				4.1.1
6900	<i>radA</i>	DNA repair protein RadA	450				OM	U				2.2.03
6910	<i>dsbE</i>	thiol:disulfide interchange protein*	166	1 [†]		signal	C	P				2.2.13
6920		conserved hypothetical protein	73				C	C				0.0.2
6930		glutaredoxin-related protein [†]	110				C	C				7.0.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
6940	<i>ligA</i>	NAD-dependent DNA ligase	674				OM	U		238 bp x 3.2 (C-terminus)	6.5.1.2	2.2.03
6950		exonuclease*	389				OM	U				7.0.0
6960		histidine kinase sensor component of a two-component regulatory system*	710	5		5 th	IM	IM				6.1.2
6970		unknown	95				C	C				7.0.0
6980		conserved hypothetical protein	280				E	U				0.0.2
6990	<i>dcd</i>	deoxycytidine triphosphate deaminase*	185				P	U			3.5.4.13	3.7.2
7000	<i>purC</i>	phosphoribosylaminoimidazole-succinocarboxamide synthase	250				C	U			6.3.2.6	3.7.1
7010	<i>hisS</i>	histidyl-tRNA synthetase	414				OM	C			6.1.1.21	2.2.01
7020		integral membrane protein*	987	2		2 th	IM	C				4.1.1
7030		disulfide oxidoreductase*	250	1 ⁺		signal	C	U				1.3.1
7040		cytochrome c oxidase assembly protein*	359	8		8 th	IM	IM				2.2.13
7050	<i>ccmA</i>	heme exporter protein A	213				C	U				1.5.7
7060		unknown	546				C	C				0.0.0
7070		membrane protein*	1373	1		1 th	OM	C		141 bp x 3.9, 198 bp x 5.2		4.1.1
7080		membrane protein*	899	1		1 th	OM	U				4.1.1
7090		membrane protein*	228	1 ⁺		1 th	C	C				4.1.1
7100		membrane protein*	250	1 ⁺		1 th	C	U				4.1.1
7110		exported protein*	182		signal	1 th	C	U				4.1.0
7120		exported protein*	204		signal	1 th	C	C				4.1.0
7130		membrane protein*	186	1 ⁺		1 th	C	C				4.1.1
7140		membrane protein*	197	1 ⁺		1 th	OM	C				4.1.1
7150		membrane protein*	142	1		1 th	C	U				4.1.1
7160		membrane protein*	172	1		1 th	C	C				4.1.1
7170		methylpurine-DNA glycosylase*	188				C	U		178 bp x 2.8 (C-terminus)	3.2.2.-	2.2.03
7180		membrane protein*	241	1 ⁺		1 th	C	U				4.1.1
7190		unknown	281				OM	C				0.0.0
7200		unknown	360				C	C				0.0.0
7220		cytidyltransferase*	228	6		6 th	IM	IM	+	181 bp x 1.9 (C-terminus)	2.7.7.41	3.6.1
7230	<i>frr</i>	ribosome recycling factor	185				C	C				2.2.10
7240	<i>pyrH</i>	uridylate kinase	244				C	U			2.7.4.-	3.3.11
7250		membrane protein*	999	1		1 th	OM	OM				4.1.1
7260	<i>mhA</i>	ribonuclease HI	146				C	U			3.1.26.4	2.1.2
7270		membrane protein*	198	1 ⁺		1 th	P	U				4.1.1
7280		membrane protein*	181	1		1 th	C	U				4.1.1
7290	<i>mfd</i>	transcription-repair coupling factor	1122				C	U				2.2.03
7300		integral membrane protein*	157	2		signal, 1 th	E	U				4.1.1

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
7310		integral membrane protein*	202	2		signal, 1 th	E	U				4.1.1
7320		integral membrane protein*	266	2		1 th	OM	U				4.1.1
7330		membrane protein*	291	1		signal, 1 th	OM	U				4.1.1
7340		membrane protein*	122	1		1 th	C	U				4.1.1
7350		membrane protein*	145	1		1 th	E	U				4.1.1
7360		membrane protein*	147	1		1 th	IM	U				4.1.1
7370		integral membrane protein*	169	2		2 th	E	U				4.1.1
7380		membrane protein*	157	1		1 th	C	C				4.1.1
7390	<i>ribE</i>	riboflavin synthase, alpha subunit*	202				E	U			2.5.1.9	3.2.13
7400		unknown	467				C	U				0.0.0
7410		zinc protease*	421				E	U			3.4.24.-	2.1.4
7420		conserved hypothetical protein	241			signal	C	U	+			0.0.2
7430	<i>secB</i>	protein-export protein SecB*	174				C	U				1.3.1
7440		conserved hypothetical protein	213			signal	C	U				0.0.1
7450		integral membrane protein*	409	12		12 th	IM	IM				4.1.1
7460	<i>tmk</i>	thymidylate kinase*	202				C	C			2.7.4.9	3.7.2
7470	<i>fabD</i>	malonyl CoA-acyl carrier protein transacylase*	320				IM	U			2.3.1.39	3.6.1
7480	<i>rpmE</i>	50S ribosomal protein L31	75				C	C				4.2.2
7490	<i>ppnK</i>	inorganic polyphosphate/ATP-NAD kinase*	263				OM	U			2.7.1.23	3.3.0
7500	<i>guaB</i>	inosine-5'-monophosphate dehydrogenase	485				C	U			1.1.1.205	3.7.1
7510		unknown	281				C	U				7.0.0
7520	<i>pdhA</i>	pyruvate dehydrogenase E1 component, alpha subunit	329				C	U			1.2.4.1	3.5.7
7530		conjugal transfer protein*	258			signal	OM	U				1.5.6
7540	<i>trxA</i>	thioredoxin 1	107				C	C				3.2.15
7550		conserved hypothetical protein	413				C	C				0.0.2
7560	<i>xseB</i>	exodeoxyribonuclease VII small subunit*	62				C	C			3.1.11.6	2.1.1
7570		NADH-ubiquinone oxidoreductase*	97				P	U			1.6.99.3	3.5.3
7580		integral membrane transport protein*	304	10		10 th	IM	IM				1.5.5
7590		conserved hypothetical protein	194				OM	U				0.0.2
7600		membrane protein*	425	1		signal	OM	U		16 bp x 1.9		4.1.1
7610	<i>gltX1</i>	glutamyl-tRNA synthetase 1	443				C	C			6.1.1.17	2.2.01
7620		integral membrane protein*	120	3		3 th	IM	U				4.1.1
7630	<i>thiG</i>	thiazole biosynthesis protein	261				C	U				3.2.14
7640		thiamin S protein [†]	74				C	U				3.2.14
7650		unknown	468				E	U				0.0.0
7660		NifU-related protein*	185				C	C				7.0.0
7661		unknown	84				C	U				7.0.0
7670	<i>rho2</i>	transcription termination factor 2	458				C	C				2.2.11
7680	<i>hslV</i>	ATP-dependent protease HslV	189				C	C			3.4.25.-	2.1.4

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
7690	<i>hslU</i>	ATP-dependent hsl protease ATP-binding subunit	488				C	C				2.1.4
7700	<i>ubiE</i>	ubiquinone/menaquinone biosynthesis methyltransferase UbiE	230				C	U			2.1.1.-	3.2.08
7710	<i>metG</i>	methionyl-tRNA synthetase	506				C	C			6.1.1.10	2.2.01
7720		aspartate kinase*	415				C	U			2.7.2.4	3.1.13
7730	<i>coxB</i>	cytochrome c oxidase subunit 2*	258	3		2 th, signal	IM	IM			1.9.3.1	3.5.3
7740	<i>coxA</i>	cytochrome c oxidase subunit 1*	518	12			IM	IM			1.9.3.1	3.5.3
7750	<i>ctaB</i>	protoheme IX farnesyltransferase*	295	9		9 th	IM	IM			2.5.1.-	3.5.1
7760		exported lipoprotein*	250		signal	signal	C	U				4.1.0
7770	<i>purD</i>	phosphoribosylamine--glycine ligase	423				C	C			6.3.4.13	3.7.1
7780		preprotein translocase subunit YajC*	122	1		1 th ,signal	C	U				1.5.5
7790		unknown	234				OM	U				0.0.0
7800		outer membrane efflux protein*	415		signal	signal	OM	OM				1.5.5
7810	<i>rplM</i>	50S ribosomal protein L13	156				P	U				4.2.2
7820	<i>rpsI</i>	30S ribosomal protein S9	153				E	U				4.2.2
7830	<i>argC</i>	N-acetyl-gamma-glutamyl-phosphate reductase	347	1		1 th	C	U			1.2.1.38	3.1.02
7840	<i>ppa</i>	inorganic pyrophosphatase	173				C	C			3.6.1.1	3.3.13
7850		unknown	209				C	U	+			7.0.0
7860		response regulator component of a two-component regulatory system*	267				C	C				6.1.2
7870		exonuclease [†]	205				C	C				7.0.0
7880	<i>dnaN</i>	DNA polymerase III, beta subunit	375				C	C			2.7.7.7	2.2.03
7890		conserved hypothetical protein	349				C	U				0.0.2
7900	<i>prsA</i>	ribose-phosphate pyrophosphokinase	318				C	U			2.7.6.1	3.7.1
7910	<i>gatC</i>	glutamyl-tRNA(Gln) amidotransferase subunit C*	114				C	U			6.3.5.-	2.2.01
7920	<i>acnA</i>	aconitate hydratase	875				OM	C			4.2.1.3	3.5.8
7930		conserved hypothetical protein	134				C	U				0.0.2
7940	<i>purK</i>	phosphoribosylaminoimidazole carboxylase ATPase subunit	359				C	U			4.1.1.21	3.7.1
7950		ATP/GTP-binding membrane protein*	735	1		1 th	OM	U				4.1.1
7960		unknown	1304			signal	OM	OM		15 bp x 2.7		7.0.0
7970		exported protein*	1710		signal	signal	OM	OM				4.1.0
7980		type IV secretion system protein [†]	790				C	C				1.5.6
7990		integral membrane protein*	125	3		signal, 2 th	IM	IM				4.1.1
8000		integral membrane protein*	112	3		signal, 2 th	IM	IM				4.1.1
8010		integral membrane protein*	118	3		signal, 2 th	IM	IM				4.1.1
8020		integral membrane protein*	124	3		signal, 2 th	IM	IM				4.1.1
8030	<i>hflK</i>	HflK protein [†]	356			1 th	C	U				7.0.0
8040	<i>hflC</i>	HflC membrane protein [†]	290	1 [†]		1 th	C	U			3.4.-.-	4.1.1
8050		exported serine protease*	476		signal	signal	OM	P			3.4.21.-	2.1.4
8060		exported protein*	204		signal	signal	OM	U				4.1.0

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
8070	<i>rnc</i>	ribonuclease III	227				C	C			3.1.26.3	2.1.2
8080	<i>ctaG</i>	cytochrome c oxidase assembly protein	177		signal	signal	P	U				2.2.13
8090		exported peptidase*	438		signal	signal	C	P			3.4.24.-	2.1.4
8100		exported M16 family peptidase*	455		signal	signal	E	U			3.4.24.-	2.1.4
8110		integral membrane protein*	224	4		4 th	IM	IM				4.1.1
8120	<i>lspA</i>	lipoprotein signal peptidase	149	3		4 th	IM	IM			3.4.23.36	2.2.06
8130	<i>ribF</i>	riboflavin kinase/FAD synthetase	306				C	C			2.7.1.26, 2.7.7.2	3.2.13
8140	<i>grxC</i>	glutaredoxin 3*	95				C	U				3.2.15
8150		methyltransferase*	280				C	U			2.1.1.-	7.0.0
8160	<i>map</i>	methionine aminopeptidase	266				C	C			3.4.11.18	2.2.10
8170		unknown	372				C	IM				0.0.0
8180		unknown	150				OM	U				0.0.0
8190		unknown	563				OM	C				0.0.0
8200	<i>sucB</i>	dihydrolipoamide succinyltransferase, E2 component of 2-oxoglutarate dehydrogenase complex	402				C	C			2.3.1.61	3.5.8
8210		transferase [†]	172				C	U				7.0.0
8220		exported D-alanyl-D-alanine carboxypeptidase*	290		signal	signal	IM	U			3.4.16.4	2.1.4
8230		integral membrane protein*	402	8		8 th	IM	IM				4.1.1
8240		conserved hypothetical protein	93				C	C				0.0.1
8250		membrane-associated zinc metalloprotease*	379	4		4 th	IM	IM			3.4.24.-	2.1.4
8260		outer membrane protein*	771		signal	signal	OM	OM				4.1.4
8270		outer membrane protein*	182		signal	signal	C	U				4.1.3
8280	<i>fabZ</i>	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase	145				C	U			4.2.1.-	3.6.1
8290	<i>purH</i>	bifunctional purine biosynthesis protein PurH	504				C	U			3.5.4.10, 2.1.2.3	3.7.1
8300	<i>pgsA</i>	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase*	185	5		6 th	IM	IM			2.7.8.5	2.2.07
8310		integral membrane protein*	433	12		12 th	IM	IM				4.1.1
8320		Surf1-like protein [†]	213	2		2 th	IM	IM				4.1.1
8330		conserved hypothetical protein	609		signal	signal	E	OM				0.0.2
8340		unknown	622				OM	U				0.0.0
8350		unknown	419			2 th	C	C				7.0.0
8360	<i>atpB</i>	ATP synthase A subunit	243	7		7 th	IM	IM			3.6.3.14	3.5.9
8370	<i>atpE</i>	ATP synthase C subunit	73	2		2 th	IM	U			3.6.3.14	3.5.9
8380	<i>atpF</i>	ATP synthase B subunit*	167	1		1 th	C	U			3.6.3.14	3.5.9
8390		membrane protein*	163	1		1 th	C	C				4.1.1
8400	<i>ftsA</i>	cell division protein FtsA	419				C	U				1.7.1
8410	<i>trkH</i>	Trk system potassium uptake protein	483	12		12 th	IM	IM				1.5.2
8420		conserved hypothetical protein	442				C	C				0.0.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
8430	<i>ftsH</i>	cell division protein FtsH	611	2 [‡]		1 th, signal	IM	IM			3.4.24.-	2.1.4
8440	<i>lgt</i>	prolipoprotein diacylglycerol transferase	259	7		7 th	IM	IM			2.4.99.-	2.2.07
8450		integral membrane protein*	560	5		5 th	OM	IM		243 bp x 5.2 (C-terminus)		4.1.1
8460		unknown	56				C	C				0.0.0
8470	<i>secD</i>	protein-export membrane protein SecD*	505	5		5 th	IM	IM				1.5.5
8480	<i>thiF</i>	adenylyltransferase ThiF*	260	1 [‡]			OM	IM			2.7.7.-	3.2.14
8490	<i>pyrE</i>	orotate phosphoribosyltransferase*	199				C	U			2.4.2.10	3.7.2
8500	<i>recA</i>	RecA protein (Recombinase A)	357				C	U				2.2.03
8510		membrane protein*	223	1			C	U			6.3.3.3	4.1.1
8520	<i>ftsY</i>	cell division protein FtsY*	309				C	C				1.7.1
8530	<i>icd</i>	isocitrate dehydrogenase [NADP]	483				C	C			1.1.1.42	3.5.8
8550	<i>recJ</i>	single-stranded-DNA-specific exonuclease RecJ*	585			2 th	C	C			3.1.-.-	2.1.1
8560		nucleic acid independent RNA polymerase*	397				C	U			2.7.7.-	2.2.11
8570	<i>ndk</i>	nucleoside diphosphate kinase	143				C	C			2.7.4.6	3.3.11
8580		transcriptional regulator [†]	201				C	U	+			6.1.2
8590	<i>map1-14</i>	outer membrane protein MAP1-14*	309		signal	signal	P	IM		6 bp x 3.7		4.1.3
8600	<i>map1-13</i>	outer membrane protein MAP1-13*	294		signal	signal	OM	IM				4.1.3
8610		exported protein*	236		signal	signal	OM	U				4.1.0
8620	<i>map1-12</i>	outer membrane protein MAP1-12*	275		signal	signal	OM	IM				4.1.3
8630	<i>map1-11</i>	outer membrane protein MAP1-11*	293		signal	signal	OM	U				4.1.3
8640	<i>map1-10</i>	outer membrane protein MAP1-10*	257		signal	signal	OM	U				4.1.3
8650	<i>map1-9</i>	outer membrane protein MAP1-9*	289		signal	signal	OM	IM				4.1.3
8660	<i>map1-8</i>	outer membrane protein MAP1-8*	282		signal	signal	OM	U				4.1.3
8670	<i>map1-7</i>	outer membrane protein MAP1-7*	283		signal	signal	OM	U				4.1.3
8680	<i>map1-6</i>	outer membrane protein MAP1-6*	295		signal	signal	OM	OM				4.1.3
8690	<i>map1-5</i>	truncated outer membrane protein MAP1-5 [†]	205				OM	U				4.1.3
8700	<i>map1-4</i>	outer membrane protein MAP1-4*	297		signal	signal	OM	IM				4.1.3
8710	<i>map1-3</i>	outer membrane protein MAP1-3*	315		signal	signal	E	U				4.1.3
8720	<i>map1-2</i>	outer membrane protein MAP1-2*	306		signal	signal	OM	U				4.1.3
8730	<i>map1-1</i>	outer membrane protein MAP1-1*	282		signal	signal	OM	U				4.1.3
8740	<i>map1</i>	major antigenic protein MAP1	290		signal	signal	E	U				4.1.4
8750	<i>map1+1</i>	outer membrane protein MAP1+1*	285			signal	OM	U				4.1.3
8760		unknown	111				C	U				0.0.0
8770		unknown	177				C	C		24 bp x 6.3		0.0.0
8780	<i>secA</i>	preprotein translocase SecA subunit	870				C	C				1.5.5
8790		unknown	143				C	U				0.0.0
8800	<i>ftsZ</i>	cell division protein FtsZ	422				C	U				1.7.1
8810		conserved hypothetical protein	135				C	U				0.0.2

Erum ID	gene name	product	length (aa)	TMHMM	SignalP	Phobius	CELLO	pSORTb	HTH	tandem repeats	EC number	class
8820		conserved hypothetical protein	157				C	U				0.0.2
8830	<i>parA</i>	chromosome partitioning protein ParA	255				OM	U				1.2.1
8840	<i>parB</i>	chromosome partitioning protein ParB	289				C	U	+			1.2.1
8850	<i>rimM</i>	16S rRNA processing protein*	172				C	U				2.2.11
8860	<i>trmD</i>	tRNA (Guanine-N(1)-)-methyltransferase	235				OM	U			2.1.1.31	2.2.11
8870	<i>rplS</i>	50S ribosomal protein L19	125				C	C				4.2.2
8880		unknown	166				C	U				7.0.0
8890	<i>thrS</i>	threonyl-tRNA synthetase	633				C	C			6.1.1.3	2.2.01
8900	<i>infC</i>	translation initiation factor IF-3	173				C	C				2.2.10
8910		conserved hypothetical protein	150				C	U				0.0.2
8920		integral membrane protein*	234	6		7 th	IM	IM				4.1.1
8930		integral membrane protein*	409	6		6 th	IM	IM				4.1.1

*probable; †possible; ‡The initial transmembrane helix could represent a possible N-terminal signal sequence.

Appendix F: Web based tools

Web based tools used in this study

Annotation

BioCyc	http://biocyc.org/ecocyc/index.shtml
KEGG	http://www.genome.jp/kegg/pathway.html
Pfam	http://pfam.sanger.ac.uk/
PROSITE	http://www.expasy.ch/prosite/
Tandem Repeats Finder	http://tandem.bu.edu/trf/trf.html

Subcellular localisation

CELLO	http://cello.life.nctu.edu.tw/
Phobius	http://phobius.cgb.ki.se/
PSORTb v.2.0	http://www.psort.org/psortb/
SignalP	http://www.cbs.dtu.dk/services/SignalP/
TMHMM2.0	http://www.cbs.dtu.dk/services/TMHMM-2.0/

Recombinant protein analysis

Protein Molecular Weight	http://www.bioinformatics.org/sms/prot_mw.html
Recombinant Protein Solubility Prediction	http://biotech.ou.edu/

Appendix G: Publications and ethics

G1: Publications

The research conducted in this study has been published in the following articles:

COLLINS, N.E., LIEBENBERG, J., DE VILLIERS, E.P., BRAYTON, K.A., LOUW, E., PRETORIUS, A., FABER, F.E., VAN HEERDEN, H., JOSEMANS, A., VAN KLEEF, M., *et al.* 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 838-843.

PRETORIUS, A., LIEBENBERG, J., LOUW, E., COLLINS, N.E. & ALLSOPP B.A. 2010. Studies of a polymorphic *E. ruminantium* gene for use as a component of a recombinant vaccine against heartwater. *Vaccine* **28**: 3531-3539.

SEBATJANE, S.I., PRETORIUS, A., LIEBENBERG, J. STEYN, H.C., & VAN KLEEF, M. 2010. *In vitro* and *in vivo* evaluation of five low molecular weight proteins of *Ehrlichia ruminantium* as potential vaccine components. *Veterinary Immunology and Immunopathology* **137**: 217-225.

Article in preparation for publication (Chapter 5):

LIEBENBERG, J., PRETORIUS, A., FABER, F.E., J. HEATH, J., COLLINS, N.E., VAN KLEEF, M. & ALLSOPP, B.A. Identification of novel potential vaccine candidates against *Ehrlichia ruminantium*. To be submitted to *Veterinary microbiology*.

G2: Ethics

The research presented in this thesis was approved by the Animal Ethics committee of the ARC-Onderstepoort Veterinary Institute and the Animal Use and Care committee of the University of Pretoria (protocol V036/06).